

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-6

```

```

Query Match          100.0%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CGVRLGC 7
        |||||
Db      1 CGVRLGC 7

```

#### RESULT 4

US-09-227-906-6

```

; Sequence 6, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-6

```

```

Query Match          100.0%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CGVRLGC 7
        |||||
Db      1 CGVRLGC 7

```

```

RESULT 5
US-09-403-089A-8
; Sequence 8, Application US/09403089A
; Patent No. 6429286
; GENERAL INFORMATION:
; APPLICANT: SUGIMURA, Kazuhisa
; TITLE OF INVENTION: Immunoregulatory Molecules and Process for Preparing
the Same
; FILE REFERENCE: 0020-4637P
; CURRENT APPLICATION NUMBER: US/09/403,089A
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP97/02540
; PRIOR FILING DATE: 1997-07-23
; PRIOR APPLICATION NUMBER: JP 9/115303
; PRIOR FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:

```

; OTHER INFORMATION: F6 amino acid sequence motif from phage random peptide library  
US-09-403-089A-8

Query Match 85.7%; Score 6; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRLGC 7  
|||  
Db 3 GVRLGC 8

RESULT 6

US-09-139-802-127

; Sequence 127, Application US/09139802  
; Patent No. 6180084  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: P-LJ 3203  
; CURRENT APPLICATION NUMBER: US/09/139,802  
; CURRENT FILING DATE: 1998-08-25  
; EARLIER APPLICATION NUMBER: 08/926,914  
; EARLIER FILING DATE: 1997-09-10  
; EARLIER APPLICATION NUMBER: 08/710,067  
; EARLIER FILING DATE: 1996-09-10  
; NUMBER OF SEQ ID NOS: 226  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 127  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Peptide  
US-09-139-802-127

Query Match 57.1%; Score 4; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRL 5  
|||  
Db 3 GVRL 6

RESULT 7

US-09-659-786-127

; Sequence 127, Application US/09659786  
; Patent No. 6491894  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata

```

; TITLE OF INVENTION:  NGR Receptor and Methods of Identifying Tumor Homing
; TITLE OF INVENTION:  Molecules That Home to Angiogenic Vasculature Using
; TITLE OF INVENTION:  Same
; FILE REFERENCE:  P-LJ 3203
; CURRENT APPLICATION NUMBER:  US/09/659,786
; CURRENT FILING DATE:  2000-09-11
; PRIOR APPLICATION NUMBER:  08/926,914
; PRIOR FILING DATE:  1997-09-10
; PRIOR APPLICATION NUMBER:  08/710,067
; PRIOR FILING DATE:  1996-09-10
; NUMBER OF SEQ ID NOS:  226
; SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 127
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-659-786-127

```

```

Query Match          57.1%;  Score 4;  DB 4;  Length 7;
Best Local Similarity 100.0%;  Pred. No. 2.5e+05;
Matches      4;  Conservative    0;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      2 GVRL 5
        ||||
Db      3 GVRL 6

```

# RESULT 8

US-08-926-914-127

```

; Sequence 127, Application US/08926914
; Patent No. 6576239
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Tumor Homing Molecules, Conjugates
; TITLE OF INVENTION: Derived Therefrom, and Methods of Using Same
; NUMBER OF SEQUENCES: 199
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,914
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```



```

; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
US-08-926-914-127

```

```

Query Match          57.1%; Score 4; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 GVRL 5
        ||||
Db      3 GVRL 6

```

# RESULT 9

US-08-318-837-30

```

; Sequence 30, Application US/08318837
; Patent No. 5981277

```

## GENERAL INFORMATION:

```

; APPLICANT: FRANSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
; APPLICANT: ANDRE; VAN HEUVERSWEYN, HUGO
; TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
; TITLE OF INVENTION: CODING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR
THERAPY OR

```

```

; TITLE OF INVENTION: IMMUNOLOGY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016

```

## COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII

```

## CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/318,837
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 800

```

## PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: PCT/EP 93/01022
; FILING DATE: 28-APR-1993

```

## PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 92.401.231.3
; FILING DATE: 30-APR-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mouse, human
; CELL LINE: PU5-1.8, THP-1
US-08-318-837-30

```

```

Query Match          57.1%; Score 4; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CGVR 4
        ||||
Db      1 CGVR 4

```

```

RESULT 10
US-08-444-818-535
; Sequence 535, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995

```

```

; ATTORNEY/AGENT INFORMATION:
;   NAME: Harbin, Alisa A.
;   REGISTRATION NUMBER: 33,895
;   REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (508)359-3876
;   TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 535:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 8 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-444-818-535

```

```

Query Match          57.1%; Score 4; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 GVRL 5
        ||||
Db      5 GVRL 8

```

```

RESULT 11
US-08-444-818-536
; Sequence 536, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
;   APPLICANT: Chien, David Y.
;   APPLICANT: Rutter, William J.
;   TITLE OF INVENTION: NANBV Diagnostics and Vaccines
;   NUMBER OF SEQUENCES: 777
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Chiron Corporation
;     STREET: 4560 Horton Street
;     CITY: Emeryville
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94608-2916
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/444,818
;     FILING DATE:
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/403,590
;     FILING DATE: 14-MAR-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Harbin, Alisa A.
;     REGISTRATION NUMBER: 33,895
;     REFERENCE/DOCKET NUMBER: 0110.002

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 536:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-444-818-536

```

```

Query Match          57.1%; Score 4; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 GVRL 5
        ||||
Db      4 GVRL 7

```

RESULT 12

US-08-444-818-537

```

; Sequence 537, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 537:

```

```
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 8 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-444-818-537
```

```
Query Match          57.1%; Score 4; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      4; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      2 GVRL 5
        ||||
Db      3 GVRL 6
```

```
RESULT 13
US-08-444-818-538
; Sequence 538, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
;   APPLICANT: Chien, David Y.
;   APPLICANT: Rutter, William J.
;   TITLE OF INVENTION: NANBV Diagnostics and Vaccines
;   NUMBER OF SEQUENCES: 777
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Chiron Corporation
;     STREET: 4560 Horton Street
;     CITY: Emeryville
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94608-2916
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/444,818
;     FILING DATE:
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/403,590
;     FILING DATE: 14-MAR-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Harbin, Alisa A.
;     REGISTRATION NUMBER: 33,895
;     REFERENCE/DOCKET NUMBER: 0110.002
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (508)359-3876
;     TELEFAX: (508)359-3885
;   INFORMATION FOR SEQ ID NO: 538:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 8 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
```

; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-444-818-538

Query Match 57.1%; Score 4; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRL 5  
|||  
Db 2 GVRL 5

RESULT 14

US-08-444-818-539

; Sequence 539, Application US/08444818

; Patent No. 6150087

; GENERAL INFORMATION:

; APPLICANT: Chien, David Y.

; APPLICANT: Rutter, William J.

; TITLE OF INVENTION: NANBV Diagnostics and Vaccines

; NUMBER OF SEQUENCES: 777

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

; COUNTRY: USA

; ZIP: 94608-2916

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/444,818

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/403,590

; FILING DATE: 14-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Harbin, Alisa A.

; REGISTRATION NUMBER: 33,895

; REFERENCE/DOCKET NUMBER: 0110.002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (508)359-3876

; TELEFAX: (508)359-3885

; INFORMATION FOR SEQ ID NO: 539:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-444-818-539

Query Match 57.1%; Score 4; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVRL 5  
    ||||  
Db 1 GVRL 4

RESULT 15

US-09-389-956-92  
; Sequence 92, Application US/09389956  
; Patent No. 6586579  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,  
; TITLE OF INVENTION: Antibodies and Methods  
; FILE REFERENCE: P-LJ 3611  
; CURRENT APPLICATION NUMBER: US/09/389,956  
; CURRENT FILING DATE: 1999-09-03  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 92  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-389-956-92

Query Match 57.1%; Score 4; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLGC 7  
    ||||  
Db 2 RLGC 5

Search completed: November 13, 2003, 10:41:55  
Job time : 7.875 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 26.9167 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-7  
Perfect score: 55  
Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,



and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	55	100.0	8	18	AAW13418	Brain homing pepti
2	55	100.0	8	21	AAB07393	Brain homing pepti
3	55	100.0	8	22	AAE11799	Phage peptide #7 t
4	55	100.0	8	23	AAU10710	Brain homing pepti
5	48	87.3	8	18	AAW13419	Brain homing pepti
6	48	87.3	8	21	AAB07394	Brain homing pepti
7	48	87.3	8	22	AAE11800	Phage peptide #8 t
8	48	87.3	8	23	AAU10711	Brain homing pepti
9	42	76.4	296	24	ABP57005	Thiobacillus ferro
10	41	74.5	23	20	AAY11429	Human 5' EST secre
11	40	72.7	52	23	ABP35141	Human ORF4114 prot
12	40	72.7	1268	22	ABB67095	Drosophila melanog
13	39	70.9	121	22	ABG59241	Human liver peptid
14	39	70.9	121	22	ABB43866	Peptide #11372 enc
15	39	70.9	121	22	ABB26791	Protein #8790 enco
16	39	70.9	121	22	AAM64849	Human brain expres
17	39	70.9	121	22	AAM77592	Human bone marrow
18	39	70.9	121	22	AAM21520	Peptide #7954 enco
19	39	70.9	121	22	AAM37781	Peptide #11818 enc
20	39	70.9	121	23	ABG46627	Human peptide enco
21	39	70.9	305	24	ABP56995	Escherichia coli D
22	39	70.9	306	22	AAG98408	Escherichia coli p
23	39	70.9	1133	22	ABG28516	Novel human diagno
24	39	70.9	1193	22	ABG18089	Novel human diagno
25	38.5	70.0	3396	22	ABB64261	Drosophila melanog
26	38	69.1	26	23	ABB96905	Omega-conopeptide
27	38	69.1	111	22	AAB94967	Human protein sequ
28	38	69.1	121	22	AAU48487	Propionibacterium
29	38	69.1	704	22	ABG19897	Novel human diagno
30	37	67.3	14	24	AAO26121	Fc region binding
31	37	67.3	292	21	AAG06385	Arabidopsis thalia
32	37	67.3	292	21	AAG53623	Arabidopsis thalia
33	37	67.3	294	22	AAM80168	Human protein SEQ
34	37	67.3	302	21	AAG06384	Arabidopsis thalia
35	37	67.3	302	21	AAG53622	Arabidopsis thalia
36	37	67.3	304	21	AAG06383	Arabidopsis thalia
37	37	67.3	304	21	AAG53621	Arabidopsis thalia
38	37	67.3	345	23	ABB92576	Herbicidally activ
39	37	67.3	745	12	AAR11809	RNA dependant RNA
40	36	65.5	20	22	AAB74174	LMW5-HL BH1 domain
41	36	65.5	21	20	AAW87835	Bcl-2 related prot
42	36	65.5	21	22	AAB74152	LMW5-HL BH1 domain
43	36	65.5	61	22	AAM94332	Human reproductive
44	36	65.5	77	21	AAB29887	Human secreted pro
45	36	65.5	77	21	AAB29888	Human secreted pro

# ALIGNMENTS

RESULT 1

AAW13418

ID AAW13418 standard; Peptide; 8 AA.

XX

AC AAW13418;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 15; Page 68; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
CC peptide that was identified using a novel method for obtaining  
CC molecules that home to a selected organ or tissue. This in vivo  
CC panning method typically involves administering a phage display  
CC library to a subject, and identifying expressed peptides which  
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
CC vascular tissue or tumour tissue. The isolated peptides (see  
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
CC labels to the selected organ/tissue (claimed) or to identify and/or  
CC isolate target molecules (claimed). The peptides can be directly  
CC identified in vivo, as compared to prior art in vitro screening  
CC methods, which require further examination to see if they maintain  
CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 55; DB 18; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 CKDWGRIC 8

Db

|||||||  
1 CKDWGRIC 8

RESULT 2

AAB07393

ID AAB07393 standard; peptide; 8 AA.

XX

AC AAB07393;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 7.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..8

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in

PT vivo panning comprises administering a library of diverse peptides

PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was

CC identified by using in vivo panning to screen a library of potential

CC organ homing molecules. The present sequence can be used to direct a

CC moiety to a the brain tissue, by linking the moiety to the present

CC sequence. Examples of potential moieties are drugs, toxins or a

CC detectable label. The present sequence contains a DXXR amino acid motif

CC (AAB12027). The DXXR motif resembles the RGD, DGR and NGR motifs that

CC bind to certain integrins.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 55; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CKDWGRIC 8  
|||  
Db 1 CKDWGRIC 8

RESULT 3

AAE11799

ID AAE11799 standard; peptide; 8 AA.

XX

AC AAE11799;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #7 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo

PT panning that selectively home to a selected organ or tissue useful for

PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 8 AA;

Query Match

100.0%; Score 55; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
      |||||||  
Db 1 CKDWGRIC 8

RESULT 4

AAU10710

ID AAU10710 standard; peptide; 8 AA.

XX

AC AAU10710;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #7 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;

KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer

CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 55; DB 23; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|||  
Db 1 CKDWGRIC 8

#### RESULT 5

AAW13419

ID AAW13419 standard; Peptide; 8 AA.

XX

AC AAW13419;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 15; Page 68; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing

CC peptide that was identified using a novel method for obtaining  
 CC molecules that home to a selected organ or tissue. This in vivo  
 CC panning method typically involves administering a phage display  
 CC library to a subject, and identifying expressed peptides which  
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
 CC vascular tissue or tumour tissue. The isolated peptides (see  
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
 CC labels to the selected organ/tissue (claimed) or to identify and/or  
 CC isolate target molecules (claimed). The peptides can be directly  
 CC identified in vivo, as compared to prior art in vitro screening  
 CC methods, which require further examination to see if they maintain  
 CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 87.3%; Score 48; DB 18; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 | | | | | | |  
 Db 1 CLDWGRIC 8

# RESULT 6

AAB07394

ID AAB07394 standard; peptide; 8 AA.

XX

AC AAB07394;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 8.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..8

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in  
PT vivo panning comprises administering a library of diverse peptides  
PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was  
CC identified by using in vivo panning to screen a library of potential  
CC organ homing molecules. The present sequence can be used to direct a  
CC moiety to a the brain tissue, by linking the moiety to the present  
CC sequence. Examples of potential moieties are drugs, toxins or a  
CC detectable label. The present sequence contains a DXXR amino acid motif  
CC (AAB12027). The DXXR motif resembles the RGD, DGR and NGR motifs that  
CC bind to certain integrins.

XX

SQ Sequence 8 AA;

Query Match 87.3%; Score 48; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8

| | | | |

Db 1 CLDWGRIC 8

#### RESULT 7

AAE11800

ID AAE11800 standard; peptide; 8 AA.

XX

AC AAE11800;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #8 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;  
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX



PT Enriched library fraction comprising molecules recovered by in vivo  
PT panning that selectively home to a selected organ or tissue useful for  
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 8 AA;

Query Match 87.3%; Score 48; DB 22; Length 8;  
Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8

| | | | |

Db 1 CLDWGRIC 8

RESULT 8

AAU10711

ID AAU10711 standard; peptide; 8 AA.

XX

AC AAU10711;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #8 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;

KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 87.3%; Score 48; DB 23; Length 8;  
Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| | | | | | |  
Db 1 CLDWGRIC 8

RESULT 9

ABP57005

ID ABP57005 standard; protein; 296 AA.

XX

AC ABP57005;

XX

DT 10-APR-2003 (first entry)

XX

DE Thiobacillus ferrooxidans D-Ala-D-Ala ligase enzyme SEQ ID NO:11.

XX

KW D-Ala-D-Ala ligase; enzyme; bacterial; structure-based drug design;  
KW protein co-ordinate data; D-Ala-D-Ala ligase inhibitor; antibacterial.

XX

OS Thiobacillus ferrooxidans.

XX

PN WO2003002063-A2.

XX

PD 09-JAN-2003.  
 XX  
 PF 28-JUN-2002; 2002WO-US20465.  
 XX  
 PR 28-JUN-2001; 2001US-301676P.  
 XX  
 PA (ESSE-) ESSENTIAL THERAPEUTICS INC.  
 PA (PLIV ) PLIVA DD ZAGREB.  
 XX  
 PI Navia MA, Ala PJ, Griffith JP, Ali JA, Faerman CH, Moe ST;  
 PI Magee AS, Connelly PR, Perola E;  
 XX  
 DR WPI; 2003-201458/19.  
 XX  
 PT Evaluating association potential of chemical entity to complex having  
 PT binding pocket defined by structural coordinates, by employing  
 PT computational unit for entity-pocket fitting operation and analyzing  
 PT the results -  
 XX  
 PS Example 8; Fig 10; 115pp; English.  
 XX  
 CC The present invention describes a method (M1) of evaluating the potential  
 CC of a chemical entity (CE) to associate with a molecule or molecular  
 CC complex comprising a binding pocket (BP) defined by specific structural  
 CC coordinates (SC) of D-Ala-D-Ala ligase (I) E. coli amino acids Lys144,  
 CC Glu180, Lys181, Leu183, Glu187, Asp257 and Glu270, by employing a  
 CC computational unit to perform a fitting operation between CE and BP  
 CC defined by SC and analysing the results of the fitting operation to  
 CC quantify the association between CE and BP. Also described is a method  
 CC (M2) for identifying a potential inhibitor of (I). M1 is useful for  
 CC evaluating the potential of a chemical entity to associate with a  
 CC molecule or molecular complex comprising a binding pocket. M2 is useful  
 CC for identifying a potential inhibitor of D-Ala-D-Ala ligase. The methods  
 CC are useful in the identification of key interaction in the active site  
 CC of the enzyme, as well as the design and optimisation of inhibitors. The  
 CC methods are also useful in the drug discovery methods, particularly for  
 CC discovering new drugs that inhibit D-Ala-D-Ala ligase, an essential  
 CC enzyme in the formation of bacterial cell walls. The present sequence  
 CC represents a D-Ala-D-Ala ligase amino acid sequence given in an example  
 CC from the present invention.  
 XX  
 SQ Sequence 296 AA;

Query Match 76.4%; Score 42; DB 24; Length 296;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
 |:||||:  
 Db 241 CRDWGRV 247

RESULT 10  
 AAY11429  
 ID AAY11429 standard; Protein; 23 AA.  
 XX  
 AC AAY11429;

XX  
 DT 21-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID No 251.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906551-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01235.  
 XX  
 PR 01-AUG-1997; 97US-0905133.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI; 1999-153781/13.  
 DR N-PSDB; AAX39495.  
 XX  
 PT New nucleic acids encoding human secreted - proteins obtained from  
 PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals  
 PT and fetal brain tissue  
 XX  
 PS Claim 34; Page 370; 434pp; English.  
 XX  
 CC AAX39440 to AAX39597 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY11374 to  
 CC AAY11531, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 XX  
 SQ Sequence 23 AA;

Query Match 74.5%; Score 41; DB 20; Length 23;  
 Best Local Similarity 62.5%; Pred. No. 4.5;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
||:| :|  
Db 7 CKNWGLLC 14

RESULT 11

ABP35141

ID ABP35141 standard; Protein; 52 AA.

XX

AC ABP35141;

XX

DT 09-JUL-2002 (first entry)

XX

DE Human ORF4114 protein, SEQ ID NO:8228.

XX

KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; haematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
KW behaviour; cancer; proliferative disorder; neurological disorder;  
KW cardiovascular disease; immune system disorder; organ transplantation;  
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;  
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX

OS Homo sapiens.

XX

PN WO200190366-A2.

XX

PD 29-NOV-2001.

XX

PF 24-MAY-2001; 2001WO-US17076.

XX

PR 24-MAY-2000; 2000US-206690P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Leach MD, Shimkets RA;

XX

DR WPI; 2002-106200/14.

DR N-PSDB; ABN79167.

XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation -

XX

PS Claim 10; Page 2302; 2508pp; English.

XX

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-

CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.

XX

SQ Sequence 52 AA;

Query Match 72.7%; Score 40; DB 23; Length 52;  
 Best Local Similarity 62.5%; Pred. No. 14;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 | | | | : |  
 Db 11 CGDWGSLC 18

# RESULT 12

ABB67095

ID ABB67095 standard; Protein; 1268 AA.

XX

AC ABB67095;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 28077.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX

OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL11198.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 28077; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1268 AA;

Query Match 72.7%; Score 40; DB 22; Length 1268;  
 Best Local Similarity 62.5%; Pred. No. 3.2e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 || || |:  
 Db 53 CKQWWRVC 60

RESULT 13  
 ABG59241  
 ID ABG59241 standard; Peptide; 121 AA.  
 XX  
 AC ABG59241;  
 XX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver peptide, SEQ ID No 37889.  
 XX

KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488898/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analysing gene expression in human adult liver -  
 XX  
 PS Claim 27; SEQ ID No 37889; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult  
 CC liver. (I) may be used for predicting, measuring and displaying gene  
 CC expression in samples derived from human adult liver. The genes  
 CC identified may be involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 CC is associated with coronary heart disease. ABG47348-ABG59930 represent  
 CC human liver single exon encoded peptides of the invention.  
 CC Note: The sequence information for this patent does not appear in the  
 CC printed specification but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 121 AA;  
  
 Query Match 70.9%; Score 39; DB 22; Length 121;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
 Qy 1 CKDWGRIC 8  
 ||| |:  
 Db 93 CKQWDRMC 100



ABB43866  
 ID ABB43866 standard; Peptide; 121 AA.  
 XX  
 AC ABB43866;  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DE Peptide #11372 encoded by human foetal liver single exon probe.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 27; SEQ ID NO 36501; 639pp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 121 AA;  
  
 Query Match 70.9%; Score 39; DB 22; Length 121;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
 Qy 1 CKDWGRIC 8  
 ||| |:  
 Db 93 CKQWDRMC 100

RESULT 15

ABB26791

ID ABB26791 standard; Protein; 121 AA.

XX

AC ABB26791;

XX

DT 23-JAN-2002 (first entry)

XX

DE Protein #8790 encoded by probe for measuring heart cell gene expression.

XX

KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

XX

OS Homo sapiens.

XX

PN WO200157274-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00666.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488899/53.

XX

PT Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX

PS Claim 15; SEQ ID No 28561; 530pp; English.

XX

CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 121 AA;

Query Match 70.9%; Score 39; DB 22; Length 121;  
Best Local Similarity 62.5%; Pred. No. 48;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| | |:  
Db 93 CKQWDRMC 100

Search completed: November 13, 2003, 09:45:27  
Job time : 27.9167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 16.5833 Seconds  
(without alignments)  
88.069 Million cell updates/sec

Title: US-09-228-866-7  
Perfect score: 55  
Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	42	76.4	296	15	US-10-186-886-11	Sequence 11, Appl
2	39	70.9	121	9	US-09-864-761-42089	Sequence 42089, A
3	39	70.9	305	15	US-10-186-886-1	Sequence 1, Appli
4	39	70.9	306	9	US-09-741-669-456	Sequence 456, App
5	38	69.1	26	11	US-09-910-082A-413	Sequence 413, App
6	37	67.3	14	12	US-10-125-869A-101	Sequence 101, App
7	36	65.5	21	15	US-10-277-693A-15	Sequence 15, Appl
8	36	65.5	61	11	US-09-764-891-2990	Sequence 2990, Ap
9	36	65.5	187	15	US-10-101-482-17	Sequence 17, Appl
10	36	65.5	296	9	US-09-815-242-11204	Sequence 11204, A
11	36	65.5	418	10	US-09-771-161A-218	Sequence 218, App
12	36	65.5	431	12	US-10-259-165-292	Sequence 292, App
13	35	63.6	14	12	US-10-125-869A-92	Sequence 92, Appl
14	35	63.6	14	12	US-10-125-869A-100	Sequence 100, App
15	35	63.6	320	15	US-10-186-886-9	Sequence 9, Appli
16	35	63.6	381	10	US-09-738-626-5815	Sequence 5815, Ap
17	35	63.6	414	15	US-10-156-761-7915	Sequence 7915, Ap
18	35	63.6	807	9	US-09-847-046-2	Sequence 2, Appli
19	35	63.6	946	9	US-09-741-669-470	Sequence 470, App
20	35	63.6	946	9	US-09-912-020-326	Sequence 326, App
21	35	63.6	1091	12	US-10-032-585-7431	Sequence 7431, Ap
22	35	63.6	1151	10	US-09-749-601A-10	Sequence 10, Appl
23	35	63.6	1151	11	US-09-912-697-33	Sequence 33, Appl
24	34	61.8	20	12	US-09-962-756-933	Sequence 933, App
25	34	61.8	20	12	US-09-962-756-961	Sequence 961, App
26	34	61.8	20	12	US-09-962-756-964	Sequence 964, App
27	34	61.8	20	12	US-09-962-756-971	Sequence 971, App
28	34	61.8	20	12	US-09-962-756-977	Sequence 977, App
29	34	61.8	20	12	US-09-962-756-981	Sequence 981, App
30	34	61.8	20	12	US-09-962-756-982	Sequence 982, App
31	34	61.8	20	12	US-09-962-756-990	Sequence 990, App
32	34	61.8	20	12	US-09-962-756-1020	Sequence 1020, Ap
33	34	61.8	20	12	US-09-962-756-1026	Sequence 1026, Ap
34	34	61.8	20	12	US-09-962-756-1028	Sequence 1028, Ap
35	34	61.8	20	12	US-09-962-756-1038	Sequence 1038, Ap
36	34	61.8	20	12	US-09-962-756-1044	Sequence 1044, Ap
37	34	61.8	20	12	US-09-962-756-1047	Sequence 1047, Ap
38	34	61.8	20	12	US-09-962-756-1051	Sequence 1051, Ap
39	34	61.8	20	12	US-09-962-756-1055	Sequence 1055, Ap
40	34	61.8	20	12	US-09-962-756-1582	Sequence 1582, Ap
41	34	61.8	20	12	US-09-962-756-1584	Sequence 1584, Ap
42	34	61.8	45	10	US-09-911-150-5	Sequence 5, Appli
43	34	61.8	162	10	US-09-747-155-177	Sequence 177, App
44	34	61.8	233	15	US-10-156-761-12450	Sequence 12450, A

## ALIGNMENTS

## RESULT 1

US-10-186-886-11

; Sequence 11, Application US/10186886

; Publication No. US20030119061A1

; GENERAL INFORMATION:

; APPLICANT: Navia, Manuel A.

; APPLICANT: Ala, Paul J.

; APPLICANT: Griffith, James P.

; APPLICANT: Ali, Janid A.

; APPLICANT: Faerman, Carlos H.

; APPLICANT: Moe, Scott T.

; APPLICANT: Magee, Andrew S.

; APPLICANT: Connolly, Patrick R.

; APPLICANT: Perola, Emanuele

; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR

; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS

ANTIBACTERIAL

; TITLE OF INVENTION: DRUGS

; FILE REFERENCE: 10283-014001

; CURRENT APPLICATION NUMBER: US/10/186,886

; CURRENT FILING DATE: 2002-06-28

; PRIOR APPLICATION NUMBER: US 60/301,676

; PRIOR FILING DATE: 2001-06-28

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 296

; TYPE: PRT

; ORGANISM: Thiobacillus ferrooxidans

US-10-186-886-11

Query Match 76.4%; Score 42; DB 15; Length 296;

Best Local Similarity 71.4%; Pred. No. 31;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7

|:||||:

Db 241 CRDWGRV 247

## RESULT 2

US-09-864-761-42089

; Sequence 42089, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES

USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 42089  
; LENGTH: 121  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP001206.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.3  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2  
; OTHER INFORMATION: EST\_HUMAN HIT: AW900149.1, EVALUATE 2.00e-03

US-09-864-761-42089

Query Match 70.9%; Score 39; DB 9; Length 121;  
Best Local Similarity 62.5%; Pred. No. 44;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CKDWGRIC 8  
|| | |:  
Db 93 CKQWDRMC 100

RESULT 3

US-10-186-886-1

; Sequence 1, Application US/10186886  
; Publication No. US20030119061A1  
; GENERAL INFORMATION:  
; APPLICANT: Navia, Manuel A.  
; APPLICANT: Ala, Paul J.  
; APPLICANT: Griffith, James P.  
; APPLICANT: Ali, Janid A.  
; APPLICANT: Faerman, Carlos H.  
; APPLICANT: Moe, Scott T.  
; APPLICANT: Magee, Andrew S.  
; APPLICANT: Connelly, Patrick R.  
; APPLICANT: Perola, Emanuele  
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR  
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS

ANTIBACTERIAL

; TITLE OF INVENTION: DRUGS  
; FILE REFERENCE: 10283-014001  
; CURRENT APPLICATION NUMBER: US/10/186,886  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/301,676  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-186-886-1

Query Match 70.9%; Score 39; DB 15; Length 305;  
Best Local Similarity 85.7%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 249 CKGWGRI 255

RESULT 4

US-09-741-669-456

; Sequence 456, Application US/09741669  
; Patent No. US20020022718A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; TITLE OF INVENTION: Genes identified as required for  
; TITLE OF INVENTION: proliferation of E. coli  
; FILE REFERENCE: ELITRA.009A  
; CURRENT APPLICATION NUMBER: US/09/741,669  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 60/173005  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 481  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 456  
; LENGTH: 306  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-669-456

Query Match 70.9%; Score 39; DB 9; Length 306;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 250 CKGWGRI 256

RESULT 5  
US-09-910-082A-413  
; Sequence 413, Application US/09910082A  
; Publication No. US20030119731A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Utah Research Foundation  
; APPLICANT: Cognetix, Inc.  
; APPLICANT: Olivera, Baldomero M.  
; APPLICANT: McIntosh, J. Michael  
; APPLICANT: Watkins, Maren  
; APPLICANT: Garrett, James E.  
; APPLICANT: Shon, Ki-Joon  
; APPLICANT: Jacobsen, Richard  
; APPLICANT: Jones, Robert M.  
; APPLICANT: Cartier, G. Edward  
; TITLE OF INVENTION: Omega-Conopeptides  
; FILE REFERENCE: 2314-241  
; CURRENT APPLICATION NUMBER: US/09/910,082A  
; CURRENT FILING DATE: 2001-07-23  
; PRIOR APPLICATION NUMBER: US 60/219,616  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 60/265,888  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 413  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 413  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Conus tulipa  
; FEATURE:  
; NAME/KEY: PEPTIDE



; LOCATION: (1)..(26)  
; OTHER INFORMATION: Xaa is Hyp  
US-09-910-082A-413

Query Match 69.1%; Score 38; DB 11; Length 26;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| || |  
Db 1 CKSWGSCX 8

RESULT 6

US-10-125-869A-101  
; Sequence 101, Application US/10125869A  
; Publication No. US20030199671A1  
; GENERAL INFORMATION:  
; APPLICANT: Rondon, Isaac Jesus  
; APPLICANT: Wu, Qi-Long  
; APPLICANT: Ley, Arthur C.  
; APPLICANT: Stochl, Mark  
; APPLICANT: Ransohoff, Thomas C.  
; APPLICANT: Potter, M. Daniel (deceased)  
; TITLE OF INVENTION: BINDING MOLECULES FOR Fc-REGION  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 3421.1006-001  
; CURRENT APPLICATION NUMBER: US/10/125,869A  
; CURRENT FILING DATE: 2002-11-19  
; PRIOR APPLICATION NUMBER: 60/284,534  
; PRIOR FILING DATE: 2001-04-18  
; NUMBER OF SEQ ID NOS: 200  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 101  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fc region binding polypeptide  
US-10-125-869A-101

Query Match 67.3%; Score 37; DB 12; Length 14;  
Best Local Similarity 62.5%; Pred. No. 14;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| || :|  
Db 4 CKRWGLMC 11

RESULT 7

US-10-277-693A-15  
; Sequence 15, Application US/10277693A  
; Publication No. US20030096367A1  
; GENERAL INFORMATION:  
; APPLICANT: Korsmeyer, Stanley J.  
; TITLE OF INVENTION: Cell Death Agonists

```

; FILE REFERENCE: 56029/36280
; CURRENT APPLICATION NUMBER: US/10/277,693A
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/379,820
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: 08/112,208
; PRIOR FILING DATE: 1993-08-26
; PRIOR APPLICATION NUMBER: 08/856,034
; PRIOR FILING DATE: 1997-05-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: Amino acid is either K (Lys) or R (Arg)
US-10-277-693A-15

```

```

Query Match          65.5%;  Score 36;  DB 15;  Length 21;
Best Local Similarity 83.3%;  Pred. No. 29;
Matches      5;  Conservative    1;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      3 DWGRIC 8
        :|||||
Db      9 NWGRIC 14

```

# RESULT 8

```

US-09-764-891-2990
; Sequence 2990, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2990
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)

```

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (60)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-2990

Query Match 65.5%; Score 36; DB 11; Length 61;  
Best Local Similarity 50.0%; Pred. No. 74;  
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| || :|  
Db 15 CHSWGNNLC 22

RESULT 9

US-10-101-482-17

; Sequence 17, Application US/10101482  
; Publication No. US20030008837A1

; GENERAL INFORMATION:

; APPLICANT: KIEFER, MICHAEL C.  
; BARR, PHILIP J.

; TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA  
; ENCODING THE PROTEINS AND METHODS OF USE THEREOF

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/101,482  
; FILING DATE: 18-Mar-2002  
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/320,157  
; FILING DATE: 07-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: LEHNHARDT, SUSAN K.  
; REGISTRATION NUMBER: 33,943  
; REFERENCE/DOCKET NUMBER: 23647-20007.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 187 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-101-482-17

Query Match 65.5%; Score 36; DB 15; Length 187;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
:|||||  
Db 96 NWGRIC 101

RESULT 10

US-09-815-242-11204  
; Sequence 11204, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11204  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-815-242-11204

Query Match 65.5%; Score 36; DB 9; Length 296;  
Best Local Similarity 37.5%; Pred. No. 2.9e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 8; Gaps 1;

Qy 1 CKDW-----GRIC 8  
|:| | |:  
Db 140 CQDWENIAQQANGRVC 155

RESULT 11

US-09-771-161A-218

; Sequence 218, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771,161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 218  
; LENGTH: 418  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-161A-218

Query Match 65.5%; Score 36; DB 10; Length 418;  
Best Local Similarity 66.7%; Pred. No. 3.9e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
| | | : : |  
Db 82 DWGKLC 87

RESULT 12

US-10-259-165-292

; Sequence 292, Application US/10259165  
; Publication No. US20030135888A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Kun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE  
SILENCING

```
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 292
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-292
```

```
Query Match          65.5%; Score 36; DB 12; Length 431;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches      5; Conservative    0; Mismatches    1; Indels      0; Gaps      0;
```

```
Qy      1 CKDWGR 6
        |||||
Db      85 CKDWAR 90
```

# RESULT 13

```
US-10-125-869A-92
; Sequence 92, Application US/10125869A
; Publication No. US20030199671A1
; GENERAL INFORMATION:
; APPLICANT: Rondon, Isaac Jesus
; APPLICANT: Wu, Qi-Long
; APPLICANT: Ley, Arthur C.
; APPLICANT: Stochl, Mark
; APPLICANT: Ransohoff, Thomas C.
; APPLICANT: Potter, M. Daniel (deceased)
; TITLE OF INVENTION: BINDING MOLECULES FOR Fc-REGION
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 3421.1006-001
; CURRENT APPLICATION NUMBER: US/10/125,869A
; CURRENT FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 60/284,534
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 200
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fc region binding polypeptide
US-10-125-869A-92
```

```
Query Match          63.6%; Score 35; DB 12; Length 14;
Best Local Similarity 62.5%; Pred. No. 30;
Matches      5; Conservative    0; Mismatches    3; Indels      0; Gaps      0;
```

Qy 1 CKDWGRIC 8  
|| || |  
Db 4 CKQWGLKC 11

RESULT 14

US-10-125-869A-100

; Sequence 100, Application US/10125869A  
; Publication No. US20030199671A1  
; GENERAL INFORMATION:  
; APPLICANT: Rondon, Isaac Jesus  
; APPLICANT: Wu, Qi-Long  
; APPLICANT: Ley, Arthur C.  
; APPLICANT: Stochl, Mark  
; APPLICANT: Ransohoff, Thomas C.  
; APPLICANT: Potter, M. Daniel (deceased)  
; TITLE OF INVENTION: BINDING MOLECULES FOR Fc-REGION  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 3421.1006-001  
; CURRENT APPLICATION NUMBER: US/10/125,869A  
; CURRENT FILING DATE: 2002-11-19  
; PRIOR APPLICATION NUMBER: 60/284,534  
; PRIOR FILING DATE: 2001-04-18  
; NUMBER OF SEQ ID NOS: 200  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 100  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fc region binding polypeptide  
US-10-125-869A-100

Query Match 63.6%; Score 35; DB 12; Length 14;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|: || :|  
Db 4 CQQWGLMC 11

RESULT 15

US-10-186-886-9

; Sequence 9, Application US/10186886  
; Publication No. US20030119061A1  
; GENERAL INFORMATION:  
; APPLICANT: Navia, Manuel A.  
; APPLICANT: Ala, Paul J.  
; APPLICANT: Griffith, James P.  
; APPLICANT: Ali, Janid A.  
; APPLICANT: Faerman, Carlos H.  
; APPLICANT: Moe, Scott T.  
; APPLICANT: Magee, Andrew S.  
; APPLICANT: Connelly, Patrick R.  
; APPLICANT: Perola, Emanuele

; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR  
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS  
ANTIBACTERIAL  
; TITLE OF INVENTION: DRUGS  
; FILE REFERENCE: 10283-014001  
; CURRENT APPLICATION NUMBER: US/10/186,886  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/301,676  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 320  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-186-886-9

Query Match 63.6%; Score 35; DB 15; Length 320;  
Best Local Similarity 57.1%; Pred. No. 4.4e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|: |||:  
Db 263 CRGWGRV 269

Search completed: November 13, 2003, 09:58:28  
Job time : 16.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 8.33333 Seconds  
(without alignments)  
92.322 Million cell updates/sec

Title: US-09-228-866-7  
Perfect score: 55  
Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries



Database : PIR\_76:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	78.2	304	2	G85068	N7-like protein [i
2	40	72.7	102	2	F72553	hypothetical prote
3	39	70.9	216	2	E69128	ribosomal protein
4	39	70.9	306	1	CEECDL	D-alanine-D-alanin
5	39	70.9	306	2	H90640	D-alanine-D-alanin
6	39	70.9	306	2	H85491	D-alanine-D-alanin
7	39	70.9	449	2	B85069	hypothetical prote
8	38	69.1	378	1	A40004	histidine decarbox
9	38	69.1	378	1	B40004	histidine decarbox
10	38	69.1	378	1	A25013	histidine decarbox
11	37	67.3	134	2	S28678	hypothetical prote
12	37	67.3	258	2	AI2234	undecaprenyl pyrop
13	37	67.3	302	2	F85068	N7 like-protein [i
14	37	67.3	345	2	T45655	1-aminocyclopropan
15	37	67.3	683	2	T40780	beta adaptin-like
16	37	67.3	723	1	RRWQTN	RNA-directed RNA p
17	37	67.3	863	2	H87556	aminopeptidase N [
18	37	67.3	882	2	AH2697	aminopeptidase N p
19	37	67.3	882	2	H97479	aminopeptidase N (
20	37	67.3	883	2	AF3417	membrane alanyl am
21	37	67.3	2212	2	T28157	erythrocyte membra
22	36	65.5	151	2	PC4164	flagellar protein
23	36	65.5	296	2	A64110	cell division inhi
24	36	65.5	322	2	H85068	N7-like protein [i
25	36	65.5	339	1	TVRTM	protein kinase (EC
26	36	65.5	418	2	A38197	protein kinase (EC
27	36	65.5	433	2	T46528	probable CDP-4-ket
28	36	65.5	437	2	E47070	CDP-4-keto-6-deoxy
29	36	65.5	437	2	S15306	CDP-4-keto-6-deoxy
30	36	65.5	437	2	AB0378	probable CDP-4-ket
31	36	65.5	437	2	AG0766	probable dehydrata
32	36	65.5	517	2	T44908	nitrite extrusion
33	36	65.5	822	2	D87325	nitrite reductase
34	36	65.5	2095	2	S29529	genome polyprotein
35	35.5	64.5	1016	2	T30553	disease resistance
36	35.5	64.5	1112	2	T10504	disease resistance
37	35	63.6	95	2	T03186	hypothetical prote
38	35	63.6	195	1	MFIVB2	matrix protein M2
39	35	63.6	247	1	WMVQ28	28K protein - pota
40	35	63.6	247	2	S03546	hypothetical prote
41	35	63.6	320	2	F82763	D-alanine-D-alanin
42	35	63.6	389	2	G82140	conserved hypothet

43	35	63.6	398	2	G82558	conserved hypothet
44	35	63.6	418	2	B69360	asparaginase (asnA
45	35	63.6	444	2	T26229	hypothetical prote

## ALIGNMENTS

### RESULT 1

G85068

N7-like protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C;Accession: G85068

R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor, Washington University in St Louis and PE Biosystems Arabidopsis Sequencing Consortium.

Nature 402, 769-777, 1999

A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A;Reference number: A85001; MUID:20083488; PMID:10617198

A;Accession: G85068

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-304 <STO>

A;Cross-references: GB:NC\_001268; NID:g7267307; PIDN:CAB81089.1; GSPDB:GN00140

C;Genetics:

A;Gene: AT4g05470

A;Map position: 4

Query Match	78.2%;	Score 43;	DB 2;	Length 304;
Best Local Similarity	75.0%;	Pred. No. 6.1;		
Matches	6;	Conservative	1;	Mismatches 1; Indels 0; Gaps 0;

Qy	1 CKDWGRIC 8
	:
Db	74 CKEWRRIC 81

### RESULT 2

F72553

hypothetical protein APE1714 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C;Accession: F72553

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.

A;Reference number: A72450; MUID:99310339; PMID:10382966

A;Accession: F72553

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-102 <KAW>  
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80715.1; PID:g5105402  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE1714  
C;Superfamily: Aeropyrum pernix hypothetical protein APE1714

Query Match 72.7%; Score 40; DB 2; Length 102;  
Best Local Similarity 62.5%; Pred. No. 7.8;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|||:|::|  
Db 21 CKDYGQLC 28

### RESULT 3

E69128  
ribosomal protein S5 - Methanobacterium thermoautotrophicum (strain Delta H)  
N;Alternate names: eukaryotic ribosomal protein S2 homolog; prokaryotic  
ribosomal protein S5 homolog  
C;Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 13-Aug-1999  
C;Accession: E69128  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;  
Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.;  
Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire,  
R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.;  
Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.;  
Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.;  
Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta  
H: functional analysis and comparative genomics.  
A;Reference number: A69000; MUID:98037514; PMID:9371463  
A;Accession: E69128  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-216 <MTH>  
A;Cross-references: GB:AE000796; GB:AE000666; NID:g2621057; PIDN:AAB84532.1;  
PID:g2621060  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MTH23  
C;Superfamily: Escherichia coli ribosomal protein S5

Query Match 70.9%; Score 39; DB 2; Length 216;  
Best Local Similarity 62.5%; Pred. No. 22;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| ||| :|  
Db 118 CGDWGCVC 125

### RESULT 4

CEECDL

D-alanine-D-alanine ligase (EC 6.3.2.4) B - Escherichia coli (strain K-12)  
N;Alternate names: alanylalanine synthetase  
C;Species: Escherichia coli  
C;Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 03-Jun-2002  
C;Accession: A30289; S40602; C37155; D64731  
R;Robinson, A.C.; Kenan, D.J.; Sweeney, J.; Donachie, W.D.  
J. Bacteriol. 167, 809-817, 1986  
A;Title: Further evidence for overlapping transcriptional units in an Escherichia coli cell envelope-cell division gene cluster: DNA sequence and transcriptional organization of the ddl ftsQ region.  
A;Reference number: A30289; MUID:86304170; PMID:3528126  
A;Accession: A30289  
A;Molecule type: DNA  
A;Residues: 1-306 <ROB>  
A;Cross-references: GB:X55034; NID:g40841; PIDN:CAA38869.1; PID:g40860  
A;Experimental source: strain K-12, substrain W3110  
R;Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobuchi, K.; Nakata, A.  
submitted to the EMBL Data Library, December 1992  
A;Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4min region.  
A;Reference number: S40531  
A;Accession: S40602  
A;Molecule type: DNA  
A;Residues: 1-306 <YUR>  
A;Cross-references: EMBL:D10483; NID:g216434; PIDN:BAA01357.1; PID:g216506  
R;Dewar, S.J.; Donachie, W.D.  
J. Bacteriol. 172, 6611-6614, 1990  
A;Title: Regulation of expression of the ftsA cell division gene by sequences in upstream genes.  
A;Reference number: A37155; MUID:91035283; PMID:2228979  
A;Accession: C37155  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 300-306 <DEW>  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: D64731  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-306 <BLAT>  
A;Cross-references: GB:AE000118; GB:U00096; NID:g1786262; PIDN:AAC73203.1; PID:g1786280; UWGP:b0092  
A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: ddlB; ddl  
A;Map position: 2 min  
A;Note: gene is located in a large cluster of genes that are involved in cell division and cell wall formation  
C;Function:  
A;Description: catalyzes ATP-driven formation of alanyl-D-alanine from 2 alanine molecules  
A;Pathway: cell wall synthesis

A;Note: two D-alanine-D-alanine ligases in *E. coli* (and *S. typhimurium*) encoded by two distinct genes; the different cellular roles and relative expression of these genes are not yet clear; however, the two enzymes display remarkably similar catalytic efficiencies and substrate specificities in spite of their differences in size and amino acid sequence

C;Superfamily: D-alanine-D-alanine ligase

C;Keywords: cell wall synthesis; dimer; ligase; magnesium

F;63-74/Region: D-alanine-D-alanine ligase motif 1

F;245-276/Region: D-alanine-D-alanine ligase motif 2

Query Match 70.9%; Score 39; DB 1; Length 306;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 250 CKGWGRI 256

#### RESULT 5

H90640

D-alanine-D-alanine ligase B [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)

C;Species: *Escherichia coli*

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: H90640

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: H90640

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-306 <HAY>

A;Cross-references: GB:BA000007; PIDN:BAB33519.1; PID:g13359552; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs0096

C;Superfamily: D-alanine-D-alanine ligase

Query Match 70.9%; Score 39; DB 2; Length 306;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 250 CKGWGRI 256

#### RESULT 6

H85491

D-alanine-D-alanine ligase B ddlB [similarity] - *Escherichia coli* (strain O157:H7, substrain EDL933)

C;Species: *Escherichia coli*  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Nov-2001  
 C;Accession: H85491  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: H85491  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-306 <STO>  
 A;Cross-references: GB:AE005174; NID:g12512798; PIDN:AAG54396.1; GSPDB:GN00145; UWGP:Z0102  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: ddlB  
 C;Superfamily: D-alanine-D-alanine ligase

Query Match 70.9%; Score 39; DB 2; Length 306;  
 Best Local Similarity 85.7%; Pred. No. 29;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
 || ||||  
 Db 250 CKGWGRI 256

#### RESULT 7

B85069  
 hypothetical protein AT4g05500 [imported] - *Arabidopsis thaliana*  
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C;Accession: B85069  
 R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor, Washington University in St Louis and PE Biosystems Arabidopsis Sequencing Consortium.  
 Nature 402, 769-777, 1999  
 A;Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.  
 A;Reference number: A85001; MUID:20083488; PMID:10617198  
 A;Accession: B85069  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-449 <STO>  
 A;Cross-references: GB:NC\_001268; NID:g7267310; PIDN:CAB81092.1; GSPDB:GN00140  
 C;Genetics:  
 A;Gene: AT4g05500  
 A;Map position: 4

Query Match 70.9%; Score 39; DB 2; Length 449;  
 Best Local Similarity 62.5%; Pred. No. 40;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy            1 CKDWGRIC 8  
              || | |:|  
Db            209 CKPWHRVC 216

RESULT 8

A40004

histidine decarboxylase (EC 4.1.1.22) - Enterobacter aerogenes

C;Species: Enterobacter aerogenes

C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 18-Jun-1999

C;Accession: A40004

R;Kamath, A.V.; Vaaler, G.L.; Snell, E.E.

J. Biol. Chem. 266, 9432-9437, 1991

A;Title: Pyridoxal phosphate-dependent histidine decarboxylases. Cloning, sequencing, and expression of genes from Klebsiella planticola and Enterobacter aerogenes and properties of the overexpressed enzymes.

A;Reference number: A40004; MUID:91236707; PMID:2033044

A;Accession: A40004

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-378 <KAM>

A;Cross-references: GB:M62745; NID:g435593; PIDN:AAA24802.1; PID:g435594

C;Superfamily: Klebsiella histidine decarboxylase

C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate

F;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match            69.1%;    Score 38;    DB 1;    Length 378;  
Best Local Similarity    62.5%;    Pred. No. 51;  
Matches        5;    Conservative        0;    Mismatches        3;    Indels        0;    Gaps        0;

Qy            1 CKDWGRIC 8  
              | ||| |  
Db            50 CGDWGEYC 57

RESULT 9

B40004

histidine decarboxylase (EC 4.1.1.22) - Klebsiella planticola

C;Species: Klebsiella planticola

C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 05-Dec-1998

C;Accession: B40004

R;Kamath, A.V.; Vaaler, G.L.; Snell, E.E.

J. Biol. Chem. 266, 9432-9437, 1991

A;Title: Pyridoxal phosphate-dependent histidine decarboxylases. Cloning, sequencing, and expression of genes from Klebsiella planticola and Enterobacter aerogenes and properties of the overexpressed enzymes.

A;Reference number: A40004; MUID:91236707; PMID:2033044

A;Accession: B40004

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-378 <KAM>

C;Superfamily: Klebsiella histidine decarboxylase

C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate

F;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 69.1%; Score 38; DB 1; Length 378;  
Best Local Similarity 62.5%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| | | | |  
Db 50 CGDWGEYC 57

RESULT 10

A25013

histidine decarboxylase (EC 4.1.1.22) - *Morganella morganii*

C;Species: *Morganella morganii*

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 18-Jun-1999

C;Accession: A25013; B26751; A26751

R;Vaaler, G.L.; Brasch, M.A.; Snell, E.E.

J. Biol. Chem. 261, 11010-11014, 1986

A;Title: Pyridoxal 5'-phosphate-dependent histidine decarboxylase. Nucleotide sequence of the hdc gene and the corresponding amino acid sequence.

A;Reference number: A25013; MUID:86278193; PMID:3015950

A;Accession: A25013

A;Molecule type: DNA

A;Residues: 1-378 <VAA>

A;Cross-references: GB:J02577; NID:g149858; PIDN:AAA25321.1; PID:g149859

A;Note: translation of initiator Met is not shown; parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing

R;Hayashi, H.; Tanase, S.; Snell, E.E.

J. Biol. Chem. 261, 11003-11009, 1986

A;Title: Pyridoxal 5'-phosphate-dependent histidine decarboxylase. Inactivation by alpha-fluoromethylhistidine and comparative sequences at the inhibitor- and coenzyme-binding sites.

A;Reference number: A92554; MUID:86278192; PMID:3733745

A;Accession: B26751

A;Molecule type: protein

A;Residues: 233-247 <HAY>

A;Note: pyridoxal phosphate site

A;Accession: A26751

A;Molecule type: protein

A;Residues: 322-334 <HA2>

A;Note: suicide inhibitor site

C;Superfamily: Klebsiella histidine decarboxylase

C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate

F;2-378/Product: histidine decarboxylase #status predicted <MAT>

F;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status experimental

F;323/Active site: Ser #status predicted

Query Match 69.1%; Score 38; DB 1; Length 378;  
Best Local Similarity 62.5%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| | | | |  
Db 50 CGDWGEYC 57



RESULT 11

S28678

hypothetical protein 1 - phage SP01

C;Species: phage SP01

C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999

C;Accession: S28678

R;Scarlatto, V.; Sayre, M.H.

Gene 114, 115-119, 1992

A;Title: Sequence of the bacteriophage SP01 gene 30.

A;Reference number: S28678; MUID:92267370; PMID:1587473

A;Accession: S28678

A;Molecule type: DNA

A;Residues: 1-134 <SCA>

A;Cross-references: EMBL:M82842; NID:g216115; PIDN:AAA32596.1; PID:g216116

C;Genetics:

A;Start codon: GTG

Query Match 67.3%; Score 37; DB 2; Length 134;  
Best Local Similarity 66.7%; Pred. No. 31;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
|||::|  
Db 116 DWGKVC 121

RESULT 12

AI2234

undecaprenyl pyrophosphate synthetase [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AI2234

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing

Cyanobacterium Anabaena sp. strain PCC 7120.

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AI2234

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-258 <KUR>

A;Cross-references: GB:BA000019; PIDN:BAB75131.1; PID:g17132565; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all3432

C;Superfamily: conserved hypothetical protein YBR002c

Query Match 67.3%; Score 37; DB 2; Length 258;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWG 5  
|||||

Db 63 CKDWG 67

RESULT 13

F85068

N7 like-protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C;Accession: F85068

R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor, Washington University in St Louis and PE Biosystems Arabidopsis Sequencing Consortium.

Nature 402, 769-777, 1999

A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A;Reference number: A85001; MUID:20083488; PMID:10617198

A;Accession: F85068

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-302 <STO>

A;Cross-references: GB:NC\_001268; NID:g7267306; PIDN:CAB81088.1; GSPDB:GN00140

C;Genetics:

A;Gene: AT4g05460

A;Map position: 4

Query Match	67.3%;	Score 37;	DB 2;	Length 302;
Best Local Similarity	50.0%;	Pred. No. 62;		
Matches	4;	Conservative	2;	Mismatches 2; Indels 0; Gaps 0;

Qy	1 CKDWGRIC 8
	:    :
Db	47 CRSWRRVC 54

RESULT 14

T45655

1-aminocyclopropane-1-carboxylic acid oxidase-like protein - Arabidopsis thaliana

N;Alternate names: protein F13I12.240

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000

C;Accession: T45655

R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.;

Artiguenave, F.; Saurin, W.; Weissenbach, J.; Mewes, H.W.; Mayer, K.F.X.;

Lemcke, K.; Schueller, C.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, November 1999

A;Reference number: Z23010

A;Accession: T45655

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-345 <CHO>

A;Cross-references: EMBL:AL133292

A;Experimental source: cultivar Columbia; BAC clone F13I12

C;Genetics:

A;Map position: 3

A;Introns: 150/2; 270/3

A;Note: F13I12.240

C;Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 67.3%; Score 37; DB 2; Length 345;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWG 5  
|||  
Db 49 CKDWG 53

RESULT 15

T40780

beta adaptin-like protein - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C;Accession: T40780

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.

submitted to the EMBL Data Library, February 1998

A;Reference number: Z21884

A;Accession: T40780

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-683 <LYN>

A;Cross-references: EMBL:AL021837; PIDN:CAA17030.1; GSPDB:GN00067;

SPDB:SPBC947.02

A;Experimental source: strain 972h-; cosmid c947

C;Genetics:

A;Gene: SPDB:SPBC947.02

A;Map position: 2

Query Match 67.3%; Score 37; DB 2; Length 683;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|:|  
Db 218 CNEWGRI 224

Search completed: November 13, 2003, 09:52:58

Job time : 9.33333 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 4.58333 Seconds  
(without alignments)  
82.083 Million cell updates/sec

Title: US-09-228-866-7  
Perfect score: 55  
Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	70.9	216	1	RS5_METTH	O26131 methanobact
2	39	70.9	305	1	DDLB_ECO57	Q8x9y6 escherichia
3	39	70.9	305	1	DDLB_ECOL6	Q8f163 escherichia
4	39	70.9	305	1	DDLB_ECOLI	P07862 escherichia
5	38	69.1	377	1	DCHS_ENTAE	P28577 enterobacte
6	38	69.1	377	1	DCHS_KLEPL	P28578 klebsiella
7	38	69.1	377	1	DCHS_MORMO	P05034 morganelle
8	37	67.3	254	1	SLBP_XENLA	P79943 xenopus lae
9	37	67.3	258	1	BDHA_ALCEU	Q9x6u2 alcaligenes
10	37	67.3	723	1	RRPO_TNVA	P22958 tobacco nec
11	37	67.3	863	1	AMPN_CAUCR	P37893 caulobacter
12	36	65.5	179	1	EAR_ASFB7	P42485 african swi
13	36	65.5	179	1	EAR_ASFE4	Q07818 african swi
14	36	65.5	179	1	EAR_ASFM2	Q07819 african swi
15	36	65.5	296	1	YFCH_HAEIN	P71373 haemophilus
16	36	65.5	339	1	KMOS_RAT	P00539 rattus norv
17	36	65.5	418	1	CDL5_HUMAN	Q14004 homo sapien
18	36	65.5	437	1	RFBH_SALTY	P26398 salmonella
19	36	65.5	2095	1	RRPL_TOSV	P37800 toscana vir
20	35	63.6	195	1	VMT2_INBSI	P08383 influenza b
21	35	63.6	247	1	V28K_PLRV1	P17518 potato leaf
22	35	63.6	247	1	V28K_PLRVW	P11621 potato leaf
23	35	63.6	320	1	DDL_XYLFA	Q9pf79 xylella fas
24	35	63.6	418	1	GATD_ARCFU	O29380 archaeoglob
25	35	63.6	523	1	RPB2_HALN1	P15352 halobacteri
26	35	63.6	555	1	SYK_METKA	Q8twp6 methanopyru
27	35	63.6	615	1	NTDO_CAEEL	Q03614 caenorhabdi
28	35	63.6	946	1	GLNE_ECOLI	P30870 escherichia
29	35	63.6	1095	1	IMB3_SCHPO	O74476 schizosacch
30	35	63.6	1122	1	RPOB_THECE	P31814 thermococcu
31	35	63.6	1195	1	RPOB_THEAC	Q03587 thermoplasm
32	35	63.6	1229	1	KPB2_FUGRU	Q9w6r1 fugu rubrip

33	34.5	62.7	474	1	MEC3_YEAST	Q02574	saccharomyc
34	34	61.8	173	1	CRBS_CYPCA	P10112	cyprinus ca
35	34	61.8	183	1	AAC1_DICDI	P14195	dictyosteli
36	34	61.8	249	1	UPPS_ANASP	P58563	anabaena sp
37	34	61.8	251	1	UPPS_ANAVA	Q9zej7	anabaena va
38	34	61.8	256	1	PTMA_CAMCO	Q45983	campylobact
39	34	61.8	378	1	FAH1_SCHPO	P78870	schizosacch
40	34	61.8	430	1	ER24_ASCIM	P78575	ascobolus i
41	34	61.8	532	1	SPER_STRPU	P16264	strongyloce
42	34	61.8	775	1	PURL_AGRT5	Q8ueb0	agrobacteri
43	34	61.8	783	1	YNR2_CAEEL	Q21988	caenorhabdi
44	34	61.8	966	1	FIB1_PETMA	P02674	petromyzon
45	34	61.8	1146	1	KMHA_DICDI	P42527	dictyosteli

# ALIGNMENTS

## RESULT 1

### RS5\_METH

ID RS5\_METTH STANDARD; PRT; 216 AA.

AC O26131;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 30S ribosomal protein S5P.

GN RPS5P OR MTH23.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

OX NCBI\_TaxID=187420;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Delta H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,

RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT deltaH: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).

CC -!- FUNCTION: With S4 and S12 plays an important role in translational

CC accuracy (By similarity).

CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S4

CC (By similarity).

CC -!- DOMAIN: The N-terminal domain interacts with the head of the 30S

CC subunit; the C-terminal domain interacts with the body and

CC contacts protein S4. The interaction surface between S4 and S5 is

CC involved in control of translational fidelity.

CC -!- SIMILARITY: Contains 1 S5 DRBM domain.

CC -!- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000796; AAB84532.1; -.  
DR PIR; E69128; E69128.  
DR HSSP; P02357; 1PKP.  
DR HAMAP; MF\_01307; -; 1.  
DR InterPro; IPR000851; Ribosomal\_S5.  
DR InterPro; IPR005324; Ribosomal\_S5\_C.  
DR InterPro; IPR005711; S5\_euk\_arch.  
DR Pfam; PF00333; Ribosomal\_S5; 1.  
DR Pfam; PF03719; Ribosomal\_S5\_C; 1.  
DR TIGRFAMs; TIGR01020; rpsE\_arch; 1.  
DR PROSITE; PS00585; RIBOSOMAL\_S5; 1.  
DR PROSITE; PS50881; S5\_DSRBD; 1.  
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
FT DOMAIN 51 114 S5 DRBM.  
SQ SEQUENCE 216 AA; 23626 MW; FC9E7D051BBB7565 CRC64;

Query Match 70.9%; Score 39; DB 1; Length 216;  
Best Local Similarity 62.5%; Pred. No. 9.3;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| | | | : |  
Db 118 CGDWGCVC 125

## RESULT 2

DDL<sub>B</sub>\_ECO57

ID DDL<sub>B</sub>\_ECO57 STANDARD; PRT; 305 AA.

AC Q8X9Y6;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE D-alanine--D-alanine ligase B (EC 6.3.2.4) (D-alanylalanine  
DE synthetase B) (D-Ala-D-Ala ligase B).

GN DDL<sub>B</sub> OR Z0102 OR ECS0096.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamosis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 CC -!- FUNCTION: Cell wall formation (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + 2 D-alanine = ADP + phosphate + D-  
 CC alanyl-D-alanine.  
 CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; second  
 CC step.  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the D-alanine--D-alanine ligase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE005186; AAG54396.1; -.  
 DR EMBL; AP002550; BAB33519.1; -.  
 DR PIR; H85491; H85491.  
 DR PIR; H90640; H90640.  
 DR HAMAP; MF\_00047; -; 1.  
 DR InterPro; IPR005905; D\_ala\_D\_ala.  
 DR InterPro; IPR000291; Dala\_lig\_Van.  
 DR Pfam; PF01820; Dala\_Dala\_ligas; 1.  
 DR TIGRFams; TIGR01205; D\_ala\_D\_alaTIGR; 1.  
 DR PROSITE; PS00843; DALA\_DALA\_LIGASE\_1; 1.  
 DR PROSITE; PS00844; DALA\_DALA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis; Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 14 14 BY SIMILARITY.  
 FT ACT\_SITE 149 149 BY SIMILARITY.  
 FT ACT\_SITE 280 280 BY SIMILARITY.  
 SQ SEQUENCE 305 AA; 32722 MW; B8C61308C79F36F1 CRC64;

Query Match 70.9%; Score 39; DB 1; Length 305;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
 |||||  
 Db 249 CKGWGRI 255

RESULT 3

DDLB\_ECOL6  
 ID DDLB\_ECOL6 STANDARD; PRT; 305 AA.  
 AC Q8FL63;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE D-alanine--D-alanine ligase B (EC 6.3.2.4) (D-alanylalanine  
 synthetase B) (D-Ala-D-Ala ligase B).  
 GN DDLB OR C0110.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -!- FUNCTION: Cell wall formation (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + 2 D-alanine = ADP + phosphate + D-  
 CC alanyl-D-alanine.  
 CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; second  
 CC step.  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the D-alanine--D-alanine ligase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE016755; AAN78608.1; -.  
 DR HAMAP; MF\_00047; -; 1.  
 DR Pfam; PF01820; Dala\_Dala\_ligas; 1.  
 DR TIGRFAMs; TIGR01205; D\_ala\_D\_alaTIGR; 1.  
 DR PROSITE; PS00843; DALA\_DALA\_LIGASE\_1; 1.  
 DR PROSITE; PS00844; DALA\_DALA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis; Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 14 14 BY SIMILARITY.  
 FT ACT\_SITE 149 149 BY SIMILARITY.  
 FT ACT\_SITE 280 280 BY SIMILARITY.  
 SQ SEQUENCE 305 AA; 32761 MW; E09D9604F7D5BF0F CRC64;

Query Match 70.9%; Score 39; DB 1; Length 305;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



Qy            1 CKDWGRI 7  
              || ||||  
Db            249 CKGWGRI 255

RESULT 4

DDL<sub>B</sub> ECOLI

ID DDL<sub>B</sub> ECOLI            STANDARD;            PRT;    305 AA.  
AC P07862;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE D-alanine--D-alanine ligase B (EC 6.3.2.4) (D-alanylalanine  
DE synthetase B) (D-Ala-D-Ala ligase B).  
GN DDL<sub>B</sub> OR DDL OR B0092.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=86304170; PubMed=3528126;  
RA Robinson A.C., Kenan D.J., Sweeney J., Donachie W.D.;  
RT "Further evidence for overlapping transcriptional units in an  
RT Escherichia coli cell envelope-cell division gene cluster: DNA  
RT sequence and transcriptional organization of the ddl ftsQ region.";  
RL J. Bacteriol. 167:809-817(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92334977; PubMed=1630901;  
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
RA Isono K., Mizobuchi K., Nakata A.;  
RT "Systematic sequencing of the Escherichia coli genome: analysis of  
RT the 0-2.4 min region.";  
RL Nucleic Acids Res. 20:3305-3308(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE OF 1-40 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=90326550; PubMed=2197603;  
RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsushashi M.;  
RT "Nucleotide sequence involving murG and murC in the mra gene cluster  
RT region of Escherichia coli.";  
RL Nucleic Acids Res. 18:4014-4014(1990).  
RN [5]

RP CHARACTERIZATION, AND PARTIAL SEQUENCE.  
 RX MEDLINE=92207163; PubMed=1554356;  
 RA Al-Bar O.A., O'Connor C.D., Giles I.G., Akhtar M.;  
 RT "D-alanine:D-alanine ligase of Escherichia coli. Expression,  
 RT purification and inhibitory studies on the cloned enzyme.";   
 RL Biochem. J. 282:747-752(1992).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=95025939; PubMed=7939684;  
 RA Fan C., Moews P.C., Walsh C.T., Knox J.R.;  
 RT "Vancomycin resistance: structure of D-alanine:D-alanine ligase at  
 RT 2.3-A resolution.";   
 RL Science 266:439-443(1994).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=97207065; PubMed=9054558;  
 RA Fan C., Park I.-S., Walsh C.T., Knox J.R.;  
 RT "D-alanine:D-alanine ligase: phosphonate and phosphinate  
 RT intermediates with wild type and the Y216F mutant.";   
 RL Biochemistry 36:2531-2538(1997).  
 CC -!- FUNCTION: Cell wall formation.  
 CC -!- CATALYTIC ACTIVITY: ATP + 2 D-alanine = ADP + phosphate + D-  
 CC alanyl-D-alanine.  
 CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; second  
 CC step.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the D-alanine--D-alanine ligase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M14029; AAA23672.1; -.  
 DR EMBL; K02668; AAA23815.1; -.  
 DR EMBL; X52644; CAA36869.1; -.  
 DR EMBL; X55034; CAA38869.1; -.  
 DR EMBL; D10483; BAB96660.1; -.  
 DR EMBL; AE000118; AAC73203.1; -.  
 DR PIR; A30289; CEECDL.  
 DR PDB; 2DLN; 01-NOV-95.  
 DR PDB; 1IOV; 12-FEB-97.  
 DR PDB; 1IOW; 12-FEB-97.  
 DR EcoGene; EG10214; ddlB.  
 DR HAMAP; MF\_00047; -; 1.  
 DR InterPro; IPR005905; D\_ala\_D\_ala.  
 DR InterPro; IPR000291; Dala\_lig\_Van.  
 DR Pfam; PF01820; Dala\_Dala\_ligas; 1.  
 DR TIGRFams; TIGR01205; D\_ala\_D\_alaTIGR; 1.  
 DR PROSITE; PS00843; DALA\_DALA\_LIGASE\_1; 1.  
 DR PROSITE; PS00844; DALA\_DALA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis; 3D-structure;  
 KW Complete proteome.

FT	INIT_MET	0	0
FT	ACT_SITE	14	14
FT	ACT_SITE	149	149
FT	ACT_SITE	280	280
FT	STRAND	3	7
FT	TURN	13	14
FT	HELIX	15	31
FT	TURN	32	33
FT	STRAND	35	39
FT	TURN	41	43
FT	HELIX	46	48
FT	TURN	49	53
FT	STRAND	54	59
FT	TURN	64	66
FT	HELIX	70	78
FT	TURN	79	79
FT	STRAND	82	82
FT	HELIX	87	94
FT	HELIX	96	105
FT	TURN	106	107
FT	STRAND	110	110
FT	STRAND	113	117
FT	HELIX	118	123
FT	TURN	127	127
FT	HELIX	128	134
FT	TURN	135	136
FT	STRAND	140	144
FT	TURN	145	146
FT	TURN	149	152
FT	STRAND	154	156
FT	HELIX	159	161
FT	HELIX	162	169
FT	TURN	170	172
FT	STRAND	175	180
FT	STRAND	186	192
FT	TURN	193	194
FT	STRAND	195	196
FT	STRAND	200	203
FT	HELIX	211	215
FT	TURN	216	216
FT	STRAND	221	223
FT	HELIX	230	247
FT	TURN	248	248
FT	STRAND	252	259
FT	TURN	261	262
FT	STRAND	265	271
FT	TURN	278	279
FT	HELIX	281	288
FT	TURN	289	290
FT	HELIX	293	302
FT	TURN	303	303
SQ	SEQUENCE	305 AA;	32708 MW; 79103A85E732A4C7 CRC64;

Query Match 70.9%; Score 39; DB 1; Length 305;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
 || |||  
 Db 249 CKGWGRI 255

RESULT 5

DCHS\_ENTAE

ID DCHS\_ENTAE STANDARD; PRT; 377 AA.  
 AC P28577;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Histidine decarboxylase (EC 4.1.1.22) (HDC).  
 GN HDC.  
 OS Enterobacter aerogenes (Aerobacter aerogenes).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Enterobacter.  
 OX NCBI\_TaxID=548;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91236707; PubMed=2033044;  
 RA Kamath A.V., Vaaler G.L., Snell E.E.;  
 RT "Pyridoxal phosphate-dependent histidine decarboxylases. Cloning,  
 RT sequencing, and expression of genes from Klebsiella planticola and  
 RT Enterobacter aerogenes and properties of the overexpressed enzymes.";  
 RL J. Biol. Chem. 266:9432-9437(1991).  
 CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).  
 CC -!- COFACTOR: Pyridoxal phosphate.  
 CC -!- SUBUNIT: Homotetramer (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,  
 CC GAD, HDC AND TYRDC).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M62745; AAA24802.1; -.  
 DR PIR; A40004; A40004.  
 DR HAMAP; MF\_00609; -; 1.  
 DR InterPro; IPR002129; Pyridoxal\_deC.  
 DR Pfam; PF00282; pyridoxal\_deC; 1.  
 DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
 KW Lyase; Decarboxylase; Pyridoxal phosphate.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).  
 SQ SEQUENCE 377 AA; 42303 MW; 4C7A3334ACA7D6AE CRC64;

Query Match 69.1%; Score 38; DB 1; Length 377;  
 Best Local Similarity 62.5%; Pred. No. 22;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 | ||| |

## RESULT 6

## DCHS\_KLEPL

ID DCHS\_KLEPL STANDARD; PRT; 377 AA.  
AC P28578; Q8KHD1; Q8KHF6;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Histidine decarboxylase (EC 4.1.1.22) (HDC).  
GN HDC.  
OS *Klebsiella planticola* (Raoultella planticola).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Raoultella.  
OX NCBI\_TaxID=575;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 43176;  
RX MEDLINE=91236707; PubMed=2033044;  
RA Kamath A.V., Vaaler G.L., Snell E.E.;  
RT "Pyridoxal phosphate-dependent histidine decarboxylases. Cloning,  
RT sequencing, and expression of genes from *Klebsiella planticola* and  
RT *Enterobacter aerogenes* and properties of the overexpressed enzymes.";  
RL J. Biol. Chem. 266:9432-9437(1991).  
RN [2]  
RP SEQUENCE OF 90-317 FROM N.A.  
RC STRAIN=19-3, 27-1, 28-1, 42-1, S8, and Y1-1;  
RX MEDLINE=22083483; PubMed=12089029;  
RA Kanki M., Yoda T., Tsukamoto T., Shibata T.;  
RT "*Klebsiella pneumoniae* produces no histamine: *Raoultella planticola*  
RT and *Raoultella ornithinolytica* strains are histamine producers.";  
RL Appl. Environ. Microbiol. 68:3462-3466(2002).  
CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).  
CC -!- COFACTOR: Pyridoxal phosphate.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- MISCELLANEOUS: This histamine-producing bacteria (HPB) causes  
CC histamine fish poisoning.  
CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,  
CC GAD, HDC AND TYRDC).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M62746; AAA25071.1; -.  
DR EMBL; AB075216; BAB97305.1; -.  
DR EMBL; AB075217; BAB97306.1; -.  
DR EMBL; AB075218; BAB97307.1; -.  
DR EMBL; AB075219; BAB97308.1; -.  
DR EMBL; AB075220; BAB97309.1; -.  
DR EMBL; AB075221; BAB97310.1; -.  
DR PIR; B40004; B40004.

DR HAMAP; MF\_00609; -; 1.  
 DR InterPro; IPR002129; Pyridoxal\_deC.  
 DR Pfam; PF00282; pyridoxal\_deC; 1.  
 DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
 KW Lyase; Decarboxylase; Pyridoxal phosphate.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).  
 FT VARIANT 147 147 A -> T (IN STRAINS 28-1 AND 42-1).  
 FT VARIANT 183 183 Q -> E (IN STRAINS 28-1 AND 42-1).  
 FT CONFLICT 155 155 R -> A (IN REF. 1).  
 SQ SEQUENCE 377 AA; 42766 MW; 131A20A0A540D25A CRC64;

Query Match 69.1%; Score 38; DB 1; Length 377;  
 Best Local Similarity 62.5%; Pred. No. 22;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 | | | | |  
 Db 49 CGDWGEYC 56

# RESULT 7

## DCHS\_MORMO

ID DCHS\_MORMO STANDARD; PRT; 377 AA.  
 AC P05034;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Histidine decarboxylase (EC 4.1.1.22) (HDC).  
 GN HDC.  
 OS *Morganella morganii* (*Proteus morganii*).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Morganella*.  
 OX NCBI\_TaxID=582;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AM-15C;  
 RX MEDLINE=86278193; PubMed=3015950;  
 RA Vaaler G.L., Brasch M.A., Snell E.E.;  
 RT "Pyridoxal 5'-phosphate-dependent histidine decarboxylase. Nucleotide  
 RT sequence of the hdc gene and the corresponding amino acid sequence.";  
 RL J. Biol. Chem. 261:11010-11014(1986).  
 RN [2]  
 RP SEQUENCE OF 232-246 AND 321-333.  
 RX MEDLINE=86278192; PubMed=3733745;  
 RA Hayashi H., Tanase S., Snell E.E.;  
 RT "Pyridoxal 5'-phosphate-dependent histidine decarboxylase.  
 RT Inactivation by alpha-fluoromethylhistidine and comparative sequences  
 RT at the inhibitor- and coenzyme-binding sites.";  
 RL J. Biol. Chem. 261:11003-11009(1986).  
 CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).  
 CC -!- COFACTOR: Pyridoxal phosphate.  
 CC -!- SUBUNIT: Homotetramer.  
 CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,  
 CC GAD, HDC AND TYRDC).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; J02577; AAA25321.1; -.  
DR PIR; A25013; A25013.  
DR HAMAP; MF\_00609; -; 1.  
DR InterPro; IPR002129; Pyridoxal\_deC.  
DR Pfam; PF00282; pyridoxal\_deC; 1.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
KW Lyase; Decarboxylase; Pyridoxal phosphate.  
FT INIT\_MET 0 0  
FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).  
FT BINDING 321 321 INHIBITOR (ALPHA-FLUOROMETHYL-  
FT HISTIDINE-PYRIDOXAL P ADDUCT).  
SQ SEQUENCE 377 AA; 42744 MW; 38AD59BA5F2BA521 CRC64;

Query Match 69.1%; Score 38; DB 1; Length 377;  
Best Local Similarity 62.5%; Pred. No. 22;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
| | | | |  
Db 49 CGDWGEYC 56

#### RESULT 8

##### SLBP\_XENLA

ID SLBP\_XENLA STANDARD; PRT; 254 AA.  
AC P79943;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Histone RNA hairpin-binding protein (Histone stem-loop binding  
DE protein 1).  
GN SLBP1 OR SLBP OR HBP.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipioidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocyte;  
RX MEDLINE=97115884; PubMed=8957003;  
RA Wang Z.-F., Whitfield M.L., Ingledue T.C. III, Dominski Z.,  
RA Marzluff W.F.;  
RT "The protein that binds the 3' end of histone mRNA: a novel RNA-  
RT binding protein required for histone pre-mRNA processing."  
RL Genes Dev. 10:3028-3040(1996).  
RN [2]  
RP PHOSPHORYLATION.  
RX MEDLINE=20387311; PubMed=10827192;  
RA Mueller B., Link J., Smythe C.;

RT "Assembly of U7 small nuclear ribonucleoprotein particle and histone  
RT RNA 3' processing in Xenopus egg extracts.";  
RL J. Biol. Chem. 275:24284-24293(2000).  
CC -!- FUNCTION: BINDS THE STEM-LOOP STRUCTURE OF REPLICATION-DEPENDENT  
CC HISTONE PRE-MRNAS AND CONTRIBUTES TO EFFICIENT 3' END PROCESSING  
CC BY STABILIZING THE COMPLEX BETWEEN HISTONE PRE-MRNA AND U7 SMALL  
CC NUCLEAR RIBONUCLEOPROTEIN (SNRNP) (BY SIMILARITY). COULD PLAY AN  
CC IMPORTANT ROLE IN TARGETING MATURE HISTONE MRNA FROM THE NUCLEUS  
CC TO THE CYTOPLASM AND TO THE TRANSLATION MACHINERY. STABILIZES  
CC MATURE HISTONE MRNA AND COULD BE INVOLVED IN CELL-CYCLE REGULATION  
CC OF HISTONE GENE EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (COILED BODIES) AND CYTOPLASMIC.  
CC -!- TISSUE SPECIFICITY: Widely expressed.  
CC -!- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN STAGE I OOCYTES, GRADUALLY  
CC INCREASING THROUGHOUT OOGENESIS. FURTHER INCREASE IS ACHIEVED  
CC DURING EARLY EMBRYOGENESIS.  
CC -!- PTM: Phosphorylated on Thr-60 during mitosis.  
CC -!- SIMILARITY: BELONGS TO THE SLBP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U75681; AAC60342.1; -.  
KW RNA-binding; mRNA processing; Nuclear protein; Phosphorylation.  
FT MOD\_RES 60 60 PHOSPHORYLATION (BY CDC2).  
FT DOMAIN 127 196 RNA-BINDING (BY SIMILARITY).  
SQ SEQUENCE 254 AA; 29726 MW; DFA0651D13D55B0C CRC64;

Query Match 67.3%; Score 37; DB 1; Length 254;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWG 5  
| | | | |  
Db 70 CKDWG 74

# RESULT 9

BDHA\_ALCEU

ID BDHA\_ALCEU STANDARD; PRT; 258 AA.  
AC Q9X6U2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE D-beta-hydroxybutyrate dehydrogenase (EC 1.1.1.30) (BDH)  
DE (3-hydroxybutyrate dehydrogenase) (3-HBDH).  
GN HBDH1.  
OS Alcaligenes eutrophus (Ralstonia eutropha).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=510;  
RN [1]



RP SEQUENCE FROM N.A.  
RC STRAIN=H16 / DSM 428 / ATCC 17699;  
RA Kim J.W., Kang D.G., Rha E.G.;  
RT "Cloning and sequencing of the gene for beta-hydroxybutyrate  
dehydrogenase from *Ralstonia eutropha*.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
CC -!- CATALYTIC ACTIVITY: (R)-3-hydroxybutanoate + NAD(+) = acetoacetate  
CC + NADH.  
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases  
CC (SDR) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF145230; AAD33952.1; -.  
DR HSSP; O70351; 1E6W.  
DR InterPro; IPR002198; ADH\_short.  
DR Pfam; PF00106; adh\_short; 1.  
DR PRINTS; PR00080; SDRFAMILY.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
KW Oxidoreductase; NAD.  
FT NP\_BIND 8 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 153 153 BY SIMILARITY.  
SQ SEQUENCE 258 AA; 27014 MW; 269A06D6CD97FAEF CRC64;

Query Match 67.3%; Score 37; DB 1; Length 258;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KDWGRI 7  
|||||  
Db 130 KDWGRI 135

RESULT 10  
RRPO\_TNVA  
ID RRPO\_TNVA STANDARD; PRT; 723 AA.  
AC P22958;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE RNA-directed RNA polymerase (EC 2.7.7.48) [Contains: 23 kDa protein].  
OS Tobacco necrosis virus (strain A) (TNV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Necrovirus.  
OX NCBI\_TaxID=12055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90320143; PubMed=2371773;  
RA Meulewaeter F., Seurinck J., van Emmelo J.;  
RT "Genome structure of tobacco necrosis virus strain A."  
RL Virology 177:699-709(1990).

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- MISCELLANEOUS: Readthrough of the terminator codon UAG occurs  
 CC between codons for Lys-202 and Gly-203.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M33002; AAA86434.2; ALT\_SEQ.  
 DR PIR; A35523; RRWQTN.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 KW Transferase; RNA-directed RNA polymerase.  
 FT CHAIN 1 202 23 kDa PROTEIN.  
 FT VARIANT 72 72 V -> A.  
 FT VARIANT 698 698 K -> R.  
 SQ SEQUENCE 723 AA; 82167 MW; DA9D142F0A3DED6D CRC64;

Query Match 67.3%; Score 37; DB 1; Length 723;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWG 5  
 |||||  
 Db 129 CKDWG 133

# RESULT 11

## AMPN\_CAUCR

ID AMPN\_CAUCR STANDARD; PRT; 863 AA.  
 AC P37893;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aminopeptidase N (EC 3.4.11.2) (Alpha-aminoacylpeptide hydrolase).  
 GN PEPN OR CC2481.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of *Caulobacter crescentus*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RN [2]  
 RP SEQUENCE OF 725-863 FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=93133840; PubMed=8421698;  
 RA Wang S.P., Sharma P.L., Schoenlein P.V., Ely B.;  
 RT "A histidine protein kinase is involved in polar organelle  
 RT development in *Caulobacter crescentus*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:630-634(1993).  
 CC -!- FUNCTION: AMINOPEPTIDASE N IS INVOLVED IN THE DEGRADATION OF  
 CC INTRACELLULAR PEPTIDES GENERATED BY PROTEIN BREAKDOWN DURING  
 CC NORMAL GROWTH AS WELL AS IN RESPONSE TO NUTRIENT STARVATION.  
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-  
 CC Xbb- from a peptide, amide or arylamide. Xaa is preferably Ala,  
 CC but may be most amino acids including Pro (slow action). When a  
 CC terminal hydrophobic residue is followed by a prolyl residue, the  
 CC two may be released as an intact Xaa-Pro dipeptide.  
 CC -!- COFACTOR: Binds 1 zinc ion (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE005917; AAK24452.1; -.  
 DR EMBL; M91449; AAA23051.1; -.  
 DR PIR; H87556; H87556.  
 DR PIR; S27532; S27532.  
 DR MEROPS; M01.005; -.  
 DR TIGR; CC2481; -.  
 DR InterPro; IPR001930; Ala\_peptase.  
 DR InterPro; IPR006025; Zn\_MTpeptdse.  
 DR Pfam; PF01433; Peptidase\_M1; 1.  
 DR PRINTS; PR00756; ALADIPTASE.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc; Complete proteome.  
 FT METAL 299 299 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 300 300 BY SIMILARITY.  
 FT METAL 303 303 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 322 322 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 383 383 PROTON DONOR (POTENTIAL).  
 SQ SEQUENCE 863 AA; 94879 MW; F04BCE19C6A5F7BD CRC64;

Query Match 67.3%; Score 37; DB 1; Length 863;  
 Best Local Similarity 50.0%; Pred. No. 68;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CKDWGRIC 8  
 | : | | : : |  
 Db 312 CRDWFQLC 319

RESULT 12

EAR\_ASFB7

ID EAR\_ASFB7 STANDARD; PRT; 179 AA.  
AC P42485;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apoptosis regulator Bcl-2 homolog precursor.  
GN A179L.  
OS African swine fever virus (strain BA71V) (ASFV).  
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
OX NCBI\_TaxID=10498;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,  
RA Rodriguez J.F., Vinuela E.;  
RT "Analysis of the complete nucleotide sequence of African swine fever  
RT virus.";  
RL Virology 208:249-278(1995).  
CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
CC CYCLE.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
CC VIRUS BHRF1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18466; AAA65271.1; -.  
DR InterPro; IPR000712; Bcl2\_BH.  
DR InterPro; IPR002475; BCL2\_family.  
DR Pfam; PF00452; Bcl-2; 1.  
DR SMART; SM00337; BCL; 1.  
DR PROSITE; PS50062; BCL2\_FAMILY; 1.  
DR PROSITE; PS01080; BH1; 1.  
DR PROSITE; PS01258; BH2; 1.  
KW Signal; Apoptosis.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 179 APOPTOSIS REGULATOR BCL-2 HOMOLOG.  
FT DOMAIN 76 95 BH1.  
FT DOMAIN 126 141 BH2.  
SQ SEQUENCE 179 AA; 21075 MW; 62CB13D82374BF35 CRC64;

Query Match 65.5%; Score 36; DB 1; Length 179;  
Best Local Similarity 83.3%; Pred. No. 25;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
:|||||

Db

83 NWGRIC 88

RESULT 13

EAR\_ASFE4

ID EAR\_ASFE4 STANDARD; PRT; 179 AA.  
AC Q07818;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apoptosis regulator Bcl-2 homolog precursor (LMH-5W).  
GN LMW5-HL.  
OS African swine fever virus (strain E-70 / isolate MS44) (ASFV).  
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
OX NCBI\_TaxID=39014;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93287262; PubMed=8389936;  
RA Neilan J.G., Lu Z., Afonso C.L., Kutish G.F., Sussman M.D., Rock D.L.;  
RT "An African swine fever virus gene with similarity to the  
RT proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1."  
RL J. Virol. 67:4391-4394(1993).  
CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
CC CYCLE.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
CC VIRUS BHRF1.  
DR InterPro; IPR000712; Bcl2\_BH.  
DR InterPro; IPR002475; BCL2\_family.  
DR Pfam; PF00452; Bcl-2; 1.  
DR SMART; SM00337; BCL; 1.  
DR PROSITE; PS50062; BCL2\_FAMILY; 1.  
DR PROSITE; PS01080; BH1; 1.  
DR PROSITE; PS01258; BH2; 1.  
KW Signal; Apoptosis.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 179 APOPTOSIS REGULATOR BCL-2 HOMOLOG.  
FT DOMAIN 76 95 BH1.  
FT DOMAIN 126 141 BH2.  
SQ SEQUENCE 179 AA; 21131 MW; 56B1C22790677BD2 CRC64;

Query Match 65.5%; Score 36; DB 1; Length 179;  
Best Local Similarity 83.3%; Pred. No. 25;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
:|||||  
Db 83 NWGRIC 88

RESULT 14

EAR\_ASFM2

ID EAR\_ASFM2 STANDARD; PRT; 179 AA.  
AC Q07819;  
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Apoptosis regulator Bcl-2 homolog precursor (LMH-5W).  
 GN LMW5-HL.  
 OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).  
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
 OX NCBI\_TaxID=10500;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93287262; PubMed=8389936;  
 RA Neilan J.G., Lu Z., Afonso C.L., Kutish G.F., Sussman M.D., Rock D.L.;  
 RT "An African swine fever virus gene with similarity to the  
 RT proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1.";  
 RL J. Virol. 67:4391-4394(1993).  
 CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
 CC CYCLE.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
 CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
 CC VIRUS BHRF1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L09548; AAA17034.1; -.  
 DR InterPro; IPR000712; Bcl2\_BH.  
 DR InterPro; IPR002475; BCL2\_family.  
 DR Pfam; PF00452; Bcl-2; 1.  
 DR SMART; SM00337; BCL; 1.  
 DR PROSITE; PS50062; BCL2\_FAMILY; 1.  
 DR PROSITE; PS01080; BH1; 1.  
 DR PROSITE; PS01258; BH2; 1.  
 KW Signal; Apoptosis.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 179 APOPTOSIS REGULATOR BCL-2 HOMOLOG.  
 FT DOMAIN 76 95 BH1.  
 FT DOMAIN 126 141 BH2.  
 SQ SEQUENCE 179 AA; 21068 MW; 0A4204D5643C66E4 CRC64;

Query Match 65.5%; Score 36; DB 1; Length 179;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DWGRIC 8  
 :|||||  
 Db 83 NWGRIC 88

RESULT 15  
 YFCH\_HAEIN  
 ID YFCH\_HAEIN STANDARD; PRT; 296 AA.

AC P71373;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein HI1208.  
 GN HI1208.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -!- SIMILARITY: BELONGS TO THE UPF0105 FAMILY. STRONG, TO E.COLI YFCH.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32800; AAC22862.1; -.  
 DR PIR; A64110; A64110.  
 DR TIGR; HI1208; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 296 AA; 33371 MW; 7AF393B7669E6C60 CRC64;

Query Match 65.5%; Score 36; DB 1; Length 296;  
 Best Local Similarity 37.5%; Pred. No. 38;  
 Matches 6; Conservative 2; Mismatches 0; Indels 8; Gaps 1;

Qy 1 CKDW-----GRIC 8  
 |:|| ||:|  
 Db 140 CQDWENIAQQANGRVC 155

Search completed: November 13, 2003, 09:46:36  
 Job time : 5.58333 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 21.0833 Seconds  
(without alignments)  
97.917 Million cell updates/sec

Title: US-09-228-866-7

Perfect score: 55

Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_23:\*

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	43	78.2	304	10	Q9M0U8	Q9m0u8 arabidopsis	
2	41	74.5	62	4	Q9P1A5	Q9pla5 homo sapien	
3	41	74.5	135	16	Q8EJF3	Q8ejf3 shewanella	
4	40	72.7	102	17	Q9YB83	Q9yb83 aeropyrum p	



5	40	72.7	115	10	Q8LKW2	Q8lkw2 medicago tr
6	40	72.7	246	10	Q9S9V8	Q9s9v8 arabidopsis
7	40	72.7	1299	5	Q9V4I9	Q9v4i9 drosophila
8	40	72.7	1299	5	Q9U5X0	Q9u5x0 drosophila
9	39	70.9	306	16	Q8FL63	Q8fl63 escherichia
10	39	70.9	449	10	Q9S9V9	Q9s9v9 arabidopsis
11	39	70.9	452	4	Q8IVP9	Q8ivp9 homo sapien
12	38.5	70.0	1551	5	Q9NGV4	Q9ngv4 drosophila
13	38.5	70.0	3396	5	Q9VM55	Q9vm55 drosophila
14	38	69.1	172	11	Q8BY83	Q8by83 mus musculu
15	38	69.1	244	11	Q9EQF4	Q9eqf4 mus musculu
16	38	69.1	275	13	O13090	O13090 pleurodeles
17	37	67.3	134	9	Q38422	Q38422 bacterioph
18	37	67.3	158	10	Q8VZB8	Q8vzb8 arabidopsis
19	37	67.3	201	12	Q83939	Q83939 olive laten
20	37	67.3	242	10	Q9FM58	Q9fm58 arabidopsis
21	37	67.3	251	16	Q8DI29	Q8di29 synechococc
22	37	67.3	258	16	Q8YRL4	Q8yrl4 anabaena sp
23	37	67.3	302	10	Q8LAJ5	Q8laj5 arabidopsis
24	37	67.3	302	10	Q9M0U9	Q9m0u9 arabidopsis
25	37	67.3	312	10	Q8W096	Q8w096 oryza sativ
26	37	67.3	345	10	Q9SD54	Q9sd54 arabidopsis
27	37	67.3	518	10	Q94HA3	Q94ha3 oryza sativ
28	37	67.3	683	3	O43079	O43079 schizosacch
29	37	67.3	723	12	Q83938	Q83938 olive laten
30	37	67.3	881	16	Q985F4	Q985f4 rhizobium l
31	37	67.3	882	16	Q8UGQ1	Q8ugq1 agrobacteri
32	37	67.3	883	16	Q8YG38	Q8yg38 brucella me
33	37	67.3	883	16	Q8G1T7	Q8g1t7 brucella su
34	37	67.3	884	16	Q92R84	Q92r84 rhizobium m
35	37	67.3	1092	3	Q9UVY2	Q9uvy2 pneumocysti
36	37	67.3	2212	5	Q94657	Q94657 plasmodium
37	36	65.5	72	12	Q8VB81	Q8vb81 white spot
38	36	65.5	151	2	Q53092	Q53092 rhodobacter
39	36	65.5	322	10	Q9M0U7	Q9m0u7 arabidopsis
40	36	65.5	340	16	Q8XHG6	Q8xhg6 clostridium
41	36	65.5	433	2	Q9ZA36	Q9za36 streptomyce
42	36	65.5	433	2	Q935Z7	Q935z7 streptomyce
43	36	65.5	434	2	Q9L4U9	Q9l4u9 streptomyce
44	36	65.5	434	2	Q9ZGC6	Q9zgc6 streptomyce
45	36	65.5	434	2	Q9L4S6	Q9l4s6 streptomyce

# ALIGNMENTS

## RESULT 1

Q9M0U8

ID Q9M0U8 PRELIMINARY; PRT; 304 AA.

AC Q9M0U8;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE N7-like protein.

GN AT4G05470.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,  
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL161503; CAB81089.1; -.  
 DR InterPro; IPR001810; F-box.  
 DR Pfam; PF00646; F-box; 1.  
 DR SMART; SM00256; FBOX; 1.  
 DR PROSITE; PS50181; FBOX; 1.  
 SQ SEQUENCE 304 AA; 34410 MW; C5EE126E80579571 CRC64;

Query Match 78.2%; Score 43; DB 10; Length 304;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 ||:| |||  
 Db 74 CKEWRRIC 81

## RESULT 2

### Q9P1A5

ID Q9P1A5 PRELIMINARY; PRT; 62 AA.  
 AC Q9P1A5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE PRO0889.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
 RA He F.;  
 RT "Functional prediction of the coding sequences of 79 new genes deduced  
 RT by analysis of cDNA clones from human fetal liver."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF119839; AAF69593.1; -.  
 SQ SEQUENCE 62 AA; 6643 MW; 478EE1DC006A36E7 CRC64;

Query Match 74.5%; Score 41; DB 4; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
 |||||  
 Db 29 DWGRIC 34

RESULT 3

Q8EJF3

ID Q8EJF3 PRELIMINARY; PRT; 135 AA.  
 AC Q8EJF3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN SO0514.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015499; AAN53595.1; -.  
 DR TIGR; SO0514; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 135 AA; 15134 MW; B54272966216A474 CRC64;

Query Match 74.5%; Score 41; DB 16; Length 135;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KDWGRIC 8  
 ||||::|  
 Db 53 KDWGQVC 59

RESULT 4

Q9YB83

ID Q9YB83 PRELIMINARY; PRT; 102 AA.  
 AC Q9YB83;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein APE1714.  
 GN APE1714.  
 OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
 OC Desulfurococcaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Res. 6:83-101(1999).  
 DR EMBL; AP000062; BAA80715.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 102 AA; 11465 MW; 3A78612D6D7A8054 CRC64;

Query Match 72.7%; Score 40; DB 17; Length 102;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 .|||:|::|  
 Db 21 CKDYGQLC 28

# RESULT 5

Q8LKW2  
 ID Q8LKW2 PRELIMINARY; PRT; 115 AA.  
 AC Q8LKW2;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Calmodulin-like protein 6b.  
 OS Medicago truncatula (Barrel medic).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
 OX NCBI\_TaxID=3880;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Jemalong;  
 RA Fedorova M., van de Mortel J.E., Matsumoto P., Town C.D.,  
 RA VandenBosch K.A., Gantt S.J., Vance C.P.;  
 RT "Genome-Wide Identification of Nodule-Specific Transcripts in the  
 RT Model Legume Medicago truncatula.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF494218; AAM81201.1; -.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; efhand; 2.  
 DR ProDom; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 SQ SEQUENCE 115 AA; 12965 MW; 83654C3307FE0DA0 CRC64;

Query Match 72.7%; Score 40; DB 10; Length 115;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| || ||  
Db 107 CKGWGFIC 114

RESULT 6

Q9S9V8

ID Q9S9V8 PRELIMINARY; PRT; 246 AA.  
AC Q9S9V8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE T1J24.1 protein.  
GN T1J24.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Ali J., Bauer C., Nguyen C., Duckels G.;  
RT "The sequence of A. thaliana T1J24.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA WashU;  
RT "The A. thaliana Genome Sequencing Project.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Waterston R.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF147263; AAD48964.1; -.  
DR InterPro; IPR001810; F-box.  
DR Pfam; PF00646; F-box; 1.  
DR SMART; SM00256; FBOX; 1.  
DR PROSITE; PS50181; FBOX; 1.  
SQ SEQUENCE 246 AA; 27438 MW; 79920E5EECF341EE CRC64;

Query Match 72.7%; Score 40; DB 10; Length 246;  
Best Local Similarity 62.5%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| || ||:  
Db 49 CKSWRRVC 56

RESULT 7

Q9V4I9

ID Q9V4I9 PRELIMINARY; PRT; 1299 AA.

AC Q9V4I9;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE CG11084 protein.

GN PK OR CG11084.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; *Drosophila*.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC  
CC IONS.  
DR EMBL; AE003842; AAF59281.2; -.  
DR HSSP; P04006; 1IML.  
DR FlyBase; FBgn0003090; pk.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00412; LIM; 2.  
DR ProDom; PD000094; LIM; 3.  
DR SMART; SM00132; LIM; 3.  
DR PROSITE; PS00478; LIM\_DOMAIN\_1; 2.  
DR PROSITE; PS50023; LIM\_DOMAIN\_2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
KW LIM domain; Metal-binding; Zinc.  
SQ SEQUENCE 1299 AA; 140721 MW; 8BFAF1F75F352485 CRC64;

Query Match 72.7%; Score 40; DB 5; Length 1299;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|| | |:

Db 53 CKQWWRVC 60

RESULT 8

Q9U5X0

ID Q9U5X0 PRELIMINARY; PRT; 1299 AA.  
AC Q9U5X0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Prickle sple isoform.  
GN PK OR CG11084.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=isogenic dp cn bw;  
RX MEDLINE=99415814; PubMed=10485852;  
RA Gubb D., Green C., Huen D., Coulson D., Johnson G., Tree D.,  
RA Collier S., Roote J.;  
RT "The balance between isoforms of the Prickle LIM domain protein is  
RT critical for planar polarity in Drosophila imaginal discs.";  
RL Genes Dev. 13:2315-2327(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=isogenic dp cn bw;  
RA Gubb D.C.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC  
CC IONS.  
DR EMBL; AJ243710; CAB57345.3; -.  
DR HSSP; P04006; 1IML.  
DR FlyBase; FBgn0003090; pk.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00412; LIM; 2.  
DR ProDom; PD000094; LIM; 3.  
DR SMART; SM00132; LIM; 3.  
DR PROSITE; PS00478; LIM\_DOMAIN\_1; 2.  
DR PROSITE; PS50023; LIM\_DOMAIN\_2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
KW LIM domain; Metal-binding; Zinc.  
SQ SEQUENCE 1299 AA; 140529 MW; 3D6D3A31717BE7DE CRC64;

Query Match 72.7%; Score 40; DB 5; Length 1299;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
||| |:  
Db 53 CKQWWRVC 60

RESULT 9



## Q8FL63

ID Q8FL63 PRELIMINARY; PRT; 306 AA.  
 AC Q8FL63;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE D-alanine--D-alanine ligase B (EC 6.3.2.4).  
 GN DDLB OR C0110.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 DR EMBL; AE016755; AAN78608.1; -.  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 306 AA; 32893 MW; 6B59AD4233475FB9 CRC64;

Query Match 70.9%; Score 39; DB 16; Length 306;  
 Best Local Similarity 85.7%; Pred. No. 56;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
 || ||||  
 Db 250 CKGWGRI 256

## RESULT 10

## Q9S9V9

ID Q9S9V9 PRELIMINARY; PRT; 449 AA.  
 AC Q9S9V9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE T1J24.2 protein (AT4G05500 protein).  
 GN T1J24.2 OR AT4G05500.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Ali J., Bauer C., Nguyen C., Duckels G.;  
 RT "The sequence of A. thaliana T1J24."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA WashU;  
 RT "The A. thaliana Genome Sequencing Project.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Waterston R.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF147263; AAD48965.1; -.  
 DR EMBL; AL161503; CAB81092.1; -.  
 DR InterPro; IPR001810; F-box.  
 DR InterPro; IPR007089; LRR\_cys.  
 DR Pfam; PF00646; F-box; 1.  
 DR SMART; SM00256; FBOX; 1.  
 DR PROSITE; PS50181; FBOX; 1.  
 DR PROSITE; PS50501; LRR\_CC; 1.  
 SQ SEQUENCE 449 AA; 51108 MW; 8EFAD4E4347718B6 CRC64;

Query Match 70.9%; Score 39; DB 10; Length 449;  
 Best Local Similarity 62.5%; Pred. No. 81;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
 || | |:  
 Db 209 CKPWHRVC 216

# RESULT 11

Q8IVP9

ID Q8IVP9 PRELIMINARY; PRT; 452 AA.  
 AC Q8IVP9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Similar to hypothetical protein FLJ32932 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC042681; AAH42681.1; -.  
 KW Hypothetical protein.

FT NON\_TER 1 1  
SQ SEQUENCE 452 AA; 51058 MW; CB71DC5BA7021312 CRC64;

Query Match 70.9%; Score 39; DB 4; Length 452;  
Best Local Similarity 85.7%; Pred. No. 81;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KDWGRIC 8  
|||  
Db 15 KDWGRRRC 21

# RESULT 12

## Q9NGV4

ID Q9NGV4 PRELIMINARY; PRT; 1551 AA.  
AC Q9NGV4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE SP1070.  
GN SP1070 OR CG9138.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Serano T.L., Pendleton J.D., Rubin G.M.;  
RT "A reverse genetic screen for genes involved in Drosophila  
RT development.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF239608; AAF63500.1; -.  
DR HSSP; P00740; 1EDM.  
DR FlyBase; FBgn0031879; SP1070.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR001438; EGF\_II.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR003410; Hyalin.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF00008; EGF; 16.  
DR Pfam; PF02494; HYR; 1.  
DR PRINTS; PR00010; EGFLOOD.  
DR SMART; SM00179; EGF\_CA; 6.  
DR SMART; SM00282; LamG; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 9.  
DR PROSITE; PS00022; EGF\_1; 15.  
DR PROSITE; PS01186; EGF\_2; 12.  
DR PROSITE; PS01187; EGF\_CA; 5.  
KW EGF-like domain.  
SQ SEQUENCE 1551 AA; 167816 MW; A97EA229E9384F31 CRC64;

Query Match 70.0%; Score 38.5; DB 5; Length 1551;  
Best Local Similarity 46.2%; Pred. No. 3.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 5; Gaps 1;

QY 1 CKDWG-----RIC 8  
| | | | : :  
Db 891 CKDWGAGGQFKVC 903

RESULT 13

Q9VM55

ID Q9VM55 PRELIMINARY; PRT; 3396 AA.  
AC Q9VM55;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG9138 protein.  
GN SP1070 OR CG9138.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -!- SIMILARITY: CONTAINS 3 CUB DOMAINS.  
 DR EMBL; AE003615; AAF52472.1; -.  
 DR HSSP; P00740; 1EDM.  
 DR FlyBase; FBgn0031879; SP1070.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000859; CUB\_domain.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000421; FA58\_C.  
 DR InterPro; IPR003410; Hyalin.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00431; CUB; 3.  
 DR Pfam; PF00008; EGF; 17.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR Pfam; PF02494; HYR; 3.  
 DR Pfam; PF00057; ldl\_recept\_a; 1.  
 DR Pfam; PF00084; sushi; 7.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR SMART; SM00032; CCP; 8.  
 DR SMART; SM00042; CUB; 3.  
 DR SMART; SM00179; EGF\_CA; 8.  
 DR SMART; SM00231; FA58C; 2.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00192; LDLA; 1.  
 DR SMART; SM00208; TNFR; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 11.  
 DR PROSITE; PS01180; CUB; 3.  
 DR PROSITE; PS00022; EGF\_1; 15.  
 DR PROSITE; PS01186; EGF\_2; 13.  
 DR PROSITE; PS01187; EGF\_CA; 7.  
 DR PROSITE; PS01285; FA58C\_1; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS50068; LDLRA\_2; 1.  
 KW EGF-like domain.  
 SQ SEQUENCE 3396 AA; 369389 MW; E618E9ACEA13E0E5 CRC64;

Query Match 70.0%; Score 38.5; DB 5; Length 3396;  
 Best Local Similarity 46.2%; Pred. No. 6.9e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 5; Gaps 1;

Qy 1 CKDWG-----RIC 8  
 ||||| :|  
 Db 2732 CKDWGAGGQFKVC 2744

RESULT 14  
 Q8BY83

ID Q8BY83 PRELIMINARY; PRT; 172 AA.  
AC Q8BY83;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Folate receptor 4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK041596; BAC30998.1; -.  
SQ SEQUENCE 172 AA; 19987 MW; C49027FEFC55BFB4 CRC64;

Query Match 69.1%; Score 38; DB 11; Length 172;  
Best Local Similarity 62.5%; Pred. No. 48;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|:| | |  
Db 130 CEDWWRAC 137

# RESULT 15

## Q9EQF4

ID Q9EQF4 PRELIMINARY; PRT; 244 AA.  
AC Q9EQF4;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Folate receptor 3 (Folate receptor 4) (Delta).  
GN FOLR4 OR FOLBP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=20564181; PubMed=11111049;  
RA Spiegelstein O., Eudy J.D., Finnell R.H.;  
RT "Identification of two putative novel folate receptor genes in humans  
RT and mouse."  
RL Gene 258:117-125(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thymus;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF250145; AAG36877.1; -.  
DR EMBL; BC028431; AAH28431.1; -.  
DR MGD; MGI:1929185; Folr4.  
DR InterPro; IPR004269; Folate\_rec.  
DR Pfam; PF03024; Folate\_rec; 1.  
KW Receptor.  
SQ SEQUENCE 244 AA; 28203 MW; 2940393EF68A52B7 CRC64;

Query Match 69.1%; Score 38; DB 11; Length 244;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CKDWGRIC 8  
|:| | | |  
Db 130 CEDWWRAC 137

Search completed: November 13, 2003, 09:51:06  
Job time : 23.0833 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 9.5 Seconds  
 (without alignments)  
 35.630 Million cell updates/sec

Title: US-09-228-866-7  
 Perfect score: 55  
 Sequence: 1 CKDWGRIC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
 Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	55	100.0	8	1 US-08-526-710-7	Sequence 7, Appli
2	55	100.0	8	3 US-08-862-855-7	Sequence 7, Appli
3	55	100.0	8	3 US-09-226-985-7	Sequence 7, Appli
4	55	100.0	8	4 US-09-227-906-7	Sequence 7, Appli
5	48	87.3	8	1 US-08-526-710-8	Sequence 8, Appli
6	48	87.3	8	3 US-08-862-855-8	Sequence 8, Appli
7	48	87.3	8	3 US-09-226-985-8	Sequence 8, Appli
8	48	87.3	8	4 US-09-227-906-8	Sequence 8, Appli
9	39	70.9	306	1 US-08-454-196-11	Sequence 11, Appl
10	39	70.9	306	2 US-08-286-819A-33	Sequence 33, Appl
11	39	70.9	306	3 US-08-980-357-33	Sequence 33, Appl



12	39	70.9	306	3	US-09-064-033-11	Sequence 11, Appl
13	39	70.9	306	4	US-09-291-046-11	Sequence 11, Appl
14	37	67.3	871	4	US-09-328-352-7076	Sequence 7076, Ap
15	36	65.5	10	2	US-08-733-505A-35	Sequence 35, Appl
16	36	65.5	10	2	US-08-706-741B-70	Sequence 70, Appl
17	36	65.5	10	2	US-08-924-695A-70	Sequence 70, Appl
18	36	65.5	20	1	US-08-248-819A-39	Sequence 39, Appl
19	36	65.5	20	2	US-08-337-646A-57	Sequence 57, Appl
20	36	65.5	20	3	US-08-927-326-57	Sequence 57, Appl
21	36	65.5	21	1	US-08-112-208C-15	Sequence 15, Appl
22	36	65.5	21	1	US-08-248-819A-17	Sequence 17, Appl
23	36	65.5	21	2	US-08-337-646A-35	Sequence 35, Appl
24	36	65.5	21	2	US-08-856-531-15	Sequence 15, Appl
25	36	65.5	21	2	US-08-856-034-15	Sequence 15, Appl
26	36	65.5	21	3	US-08-927-326-35	Sequence 35, Appl
27	36	65.5	21	4	US-09-379-820A-15	Sequence 15, Appl
28	36	65.5	67	1	US-08-321-071A-11	Sequence 11, Appl
29	36	65.5	179	1	US-08-607-269-27	Sequence 27, Appl
30	36	65.5	179	5	PCT-US95-04600-27	Sequence 27, Appl
31	36	65.5	187	1	US-08-471-058-17	Sequence 17, Appl
32	36	65.5	187	3	US-08-471-057-17	Sequence 17, Appl
33	36	65.5	187	4	US-08-470-865-17	Sequence 17, Appl
34	36	65.5	436	2	US-08-576-626A-47	Sequence 47, Appl
35	36	65.5	462	3	US-09-036-987A-18	Sequence 18, Appl
36	36	65.5	462	3	US-09-370-700-18	Sequence 18, Appl
37	36	65.5	462	4	US-09-603-207-18	Sequence 18, Appl
38	36	65.5	1037	4	US-09-428-711A-21	Sequence 21, Appl
39	35.5	64.5	799	3	US-09-180-439-6	Sequence 6, Appli
40	35.5	64.5	968	3	US-09-180-439-3	Sequence 3, Appli
41	35.5	64.5	968	3	US-09-180-439-4	Sequence 4, Appli
42	35.5	64.5	1016	3	US-09-180-439-8	Sequence 8, Appli
43	35.5	64.5	1112	3	US-09-353-585-2	Sequence 2, Appli
44	35.5	64.5	1112	3	US-09-353-585-3	Sequence 3, Appli
45	35	63.6	388	4	US-09-252-991A-31265	Sequence 31265, A

#### ALIGNMENTS

#### RESULT 1

US-08-526-710-7

; Sequence 7, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-7

```

```

Query Match          100.0%; Score 55; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CKDWGRIC 8
        |||||
Db      1 CKDWGRIC 8

```

## RESULT 2

US-08-862-855-7

```

; Sequence 7, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-7

```

```

Query Match          100.0%; Score 55; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CKDWGRIC 8
        |||||
Db      1 CKDWGRIC 8

```

```

RESULT 3
US-09-226-985-7
; Sequence 7, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-7

```

```

Query Match          100.0%; Score 55; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CKDWGRIC 8
        |||||
Db      1 CKDWGRIC 8

```

#### RESULT 4

US-09-227-906-7

```

; Sequence 7, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-7

```

```

Query Match          100.0%; Score 55; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CKDWGRIC 8
        |||||
Db      1 CKDWGRIC 8

```

# RESULT 5

US-08-526-710-8

```

; Sequence 8, Application US/08526710
; Patent No. 5622699
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-8

```

```

Query Match          87.3%; Score 48; DB 1; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CKDWGRIC 8
        | |||||
Db      1 CLDWGRIC 8

```

# RESULT 6

US-08-862-855-8

; Sequence 8, Application US/08862855

; Patent No. 6068829

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell & Flores LLP

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/862,855

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/526,710

; FILING DATE: 11-SEP-1995

; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-8

```

```

Query Match          87.3%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CKDWGRIC 8
        | |||||
Db      1 CLDWGRIC 8

```

# RESULT 7

US-09-226-985-8

```

; Sequence 8, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-8

```

```

Query Match          87.3%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 CKDWGRIC 8
        | |||||
Db      1 CLDWGRIC 8

```

# RESULT 8

US-09-227-906-8

```

; Sequence 8, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```



RESULT 12

S14958

alpha-amylase (EC 3.2.1.1) - rice

C;Species: Oryza sativa (rice)

C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999

C;Accession: S14958

R;Sutliff, T.D.; Huang, N.; Litts, J.C.; Rodriguez, R.L.

Plant Mol. Biol. 16, 579-591, 1991

A;Title: Characterization of an alpha-amylase multigene cluster in rice.

A;Reference number: S14956; MUID:91329692; PMID:1714318

A;Accession: S14958

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-440 <SUT>

A;Cross-references: EMBL:X56336; NID:g20334; PIDN:CAA39776.1; PID:g20335

C;Genetics:

A;Introns: 33/3; 78/1; 346/3

C;Function:

A;Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A;Pathway: glycogen/starch degradation

C;Superfamily: wheat alpha-amylase; alpha-amylase core homology

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;174-318/Domain: alpha-amylase core homology <AMY>

F;207,232,315/Active site: Asp, Glu, Asp #status predicted

Query Match 68.5%; Score 37; DB 2; Length 440;

Best Local Similarity 70.0%; Pred. No. 77;

Matches 7; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy 1 CLDWG--RIC 8

|||||

Db 143 CLDWGPSMIC 152

RESULT 13

B69360

asparaginase (asnA) homolog - Archaeoglobus fulgidus

C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 02-Aug-2002

C;Accession: B69360

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.;

Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;

Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: B69360

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA  
A;Residues: 1-418 <KLE>  
A;Cross-references: GB:AE001043; GB:AE000782; NID:g2689366; PIDN:AAB90360.1;  
PID:g2649722; TIGR:AF0882  
C;Superfamily: asparaginase

Query Match 67.6%; Score 36.5; DB 2; Length 418;  
Best Local Similarity 75.0%; Pred. No. 90;  
Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CLDWGRIC 8  
|| |||:  
Db 352 CL-WGRVC 358

#### RESULT 14

G85068  
N7-like protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: G85068  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor, Washington University in St Louis and PE Biosystems Arabidopsis Sequencing Consortium.  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: G85068  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-304 <STO>  
A;Cross-references: GB:NC\_001268; NID:g7267307; PIDN:CAB81089.1; GSPDB:GN00140  
C;Genetics:  
A;Gene: AT4g05470  
A;Map position: 4

Query Match 66.7%; Score 36; DB 2; Length 304;  
Best Local Similarity 62.5%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| :| |||  
Db 74 CKEWRRIC 81

#### RESULT 15

A40004  
histidine decarboxylase (EC 4.1.1.22) - Enterobacter aerogenes  
C;Species: Enterobacter aerogenes  
C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 18-Jun-1999  
C;Accession: A40004  
R;Kamath, A.V.; Vaaler, G.L.; Snell, E.E.  
J. Biol. Chem. 266, 9432-9437, 1991  
A;Title: Pyridoxal phosphate-dependent histidine decarboxylases. Cloning, sequencing, and expression of genes from Klebsiella planticola and Enterobacter aerogenes and properties of the overexpressed enzymes.

A;Reference number: A40004; MUID:91236707; PMID:2033044  
A;Accession: A40004  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-378 <KAM>  
A;Cross-references: GB:M62745; NID:g435593; PIDN:AAA24802.1; PID:g435594  
C;Superfamily: Klebsiella histidine decarboxylase  
C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate  
F;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 66.7%; Score 36; DB 1; Length 378;  
Best Local Similarity 62.5%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| | | |  
Db 50 CGDWGEYC 57

Search completed: November 13, 2003, 09:52:59  
Job time : 9.33333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 4.58333 Seconds  
(without alignments)  
82.083 Million cell updates/sec

Title: US-09-228-866-8  
Perfect score: 54  
Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	% Query		Match	Length	DB	ID	Description
	Score	Match					
1	41	75.9	247	1	V28K_PLRV1	P17518 potato leaf	
2	41	75.9	247	1	V28K_PLRVW	P11621 potato leaf	
3	39	72.2	783	1	YNR2_CAEEL	Q21988 caenorhabdi	
4	38	70.4	138	1	ACP1_SPIOL	P07854 spinacia ol	
5	38	70.4	179	1	EAR_ASFB7	P42485 african swi	
6	38	70.4	179	1	EAR_ASFE4	Q07818 african swi	
7	38	70.4	179	1	EAR_ASFM2	Q07819 african swi	
8	38	70.4	310	1	ARCC_RHIET	O31019 rhizobium e	
9	38	70.4	339	1	KMOS_RAT	P00539 rattus norv	
10	38	70.4	418	1	CDL5_HUMAN	Q14004 homo sapien	
11	37	68.5	216	1	RS5_METTH	O26131 methanobact	
12	37	68.5	440	1	AM3A_ORYSA	P27932 oryza sativ	
13	36.5	67.6	418	1	GATD_ARCFU	O29380 archaeoglob	
14		66.7	377	1	DCHS_ENTAE	P28577 enterobacte	
15		66.7	377	1	DCHS_KLEPL	P28578 klebsiella	
16	36	66.7	377	1	DCHS_MORMO	P05034 morganella	
17	36	66.7	485	1	SAHH_WHEAT	P32112 triticum ae	
18	36	66.7	1507	1	PRDF_HUMAN	P57071 homo sapien	
19	35	64.8	173	1	CRBS_CYPCA	P10112 cyprinus ca	
20	35	64.8	523	1	RPB2_HALN1	P15352 halobacteri	
21	35	64.8	555	1	SYK_METKA	Q8twp6 methanopyru	
22	35	64.8	615	1	NTDO_CAEEL	Q03614 caenorhabdi	
23	35	64.8	665	1	PDI2_HUMAN	Q9y2j8 homo sapien	
24	35	64.8	665	1	PDI2_RAT	P20717 rattus norv	
25	35	64.8	673	1	PDI2_MOUSE	Q08642 mus musculu	
26	35	64.8	1095	1	IMB3_SCHPO	O74476 schizosacch	
27	35	64.8	1122	1	RPOB_THECE	P31814 thermococcu	
28	35	64.8	1195	1	RPOB_THEAC	Q03587 thermoplasm	
29	35	64.8	4660	1	LRP2_RAT	P98158 rattus norv	
30	34.5	63.9	662	1	DNLJ_CHLPN	Q9z934 chlamydia p	
31		63.0	76	1	CXO1_CONTE	Q9xzk8 conus texti	
32		63.0	157	1	SMP1_HUMAN	O95807 homo sapien	
33	34	63.0	157	1	SMP1_MOUSE	Q9cx11 mus musculu	
34	34	63.0	255	1	UNG_EBV	P12888 epstein-bar	
35	34	63.0	349	1	KMOS_CHICK	P10741 gallus gall	
36	34	63.0	437	1	RFBH_SALTY	P26398 salmonella	
37	34	63.0	497	1	DHAB_SPIOL	P17202 spinacia ol	
38	34	63.0	500	1	DHAB_BETVU	P28237 beta vulgar	
39	34	63.0	502	1	DHAB_ATRHO	P42757 atriplex ho	
40	34	63.0	532	1	SPER_STRPU	P16264 strongyloce	
41	34	63.0	571	1	DFA1_SYNEL	Q8djy2 synechococc	
42	34	63.0	576	1	DFA1_ANASP	Q8ynw5 anabaena sp	
43	34	63.0	578	1	DFA2_SYNY3	P72723 synechocyst	
44	34	63.0	579	1	DFA2_ANASP	Q8z0c0 anabaena sp	
45	34	63.0	867	1	ENV_HV1J3	P12489 human immun	

## ALIGNMENTS

RESULT 1  
V28K\_PLRV1

ID V28K\_PLRV1 STANDARD; PRT; 247 AA.  
AC P17518;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 28 kDa protein (ORF 1).  
OS Potato leafroll virus (strain 1) (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89279282; PubMed=2732710;  
RA Mayo M.A., Robinson D.J., Jolly C.A., Hyman L.;  
RT "Nucleotide sequence of potato leafroll luteovirus RNA."  
RL J. Gen. Virol. 70:1037-1051(1989).  
CC -!- SIMILARITY: ORF1 SHOWS NO SIMILARITY WITH ANY OF THE DIFFERENT  
CC ORFS OF BARLEY YELLOW DWARF VIRUS AND BEET WESTERN YELLOWS VIRUS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00530; BAA00416.1; -.  
DR PIR; JA0117; WMVQ28.  
SQ SEQUENCE 247 AA; 28130 MW; 02E900959E8F0CEE CRC64;

Query Match 75.9%; Score 41; DB 1; Length 247;  
Best Local Similarity 62.5%; Pred. No. 4.1;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| | : | | : |  
Db 84 CLEWGLLC 91

## RESULT 2

V28K\_PLRVW

ID V28K\_PLRVW STANDARD; PRT; 247 AA.  
AC P11621;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DE 28 kDa protein (ORF 1).  
OS Potato leafroll virus (strain Wageningen) (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12048;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89171329; PubMed=2466700;  
RA van der Wilk F., Huisman M.J., Cornelissen B.J.C., Huttinga H.,  
RA Goldbach R.W.;

RT "Nucleotide sequence and organization of potato leafroll virus  
 RT genomic RNA.";  
 RL FEBS Lett. 245:51-56(1989).  
 CC -!- SIMILARITY: ORF1 SHOWS NO SIMILARITY WITH ANY OF THE DIFFERENT  
 CC ORFS OF BARLEY YELLOW DWARF VIRUS AND BEET WESTERN YELLOWS VIRUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y07496; CAA68794.1; -.  
 DR PIR; S03546; S03546.  
 SQ SEQUENCE 247 AA; 28128 MW; D730FB2728482D56 CRC64;

Query Match 75.9%; Score 41; DB 1; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 4.1;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:|| :|  
 Db 84 CLEWGLLC 91

### RESULT 3

YNR2\_CAEEL

ID YNR2\_CAEEL STANDARD; PRT; 783 AA.  
 AC Q21988;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein R13G10.2 in chromosome III.  
 GN R13G10.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Gardner A.E.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RA Durbin R.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- COFACTOR: FAD (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE FLAVIN MONOAMINE OXIDASE FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z35602; CAA84671.2; -.  
 DR WormPep; R13G10.2; CE25088.  
 DR InterPro; IPR002937; Amino\_oxidase.  
 DR InterPro; IPR000960; Flav\_cont\_mnoxgn.  
 DR Pfam; PF01593; Amino\_oxidase; 1.  
 DR PRINTS; PR00370; FMOXYGENASE.  
 KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD.  
 FT NP\_BIND 311 366 FAD (ADP PART) (POTENTIAL).  
 SQ SEQUENCE 783 AA; 88799 MW; 8D087E96464DC908 CRC64;

Query Match 72.2%; Score 39; DB 1; Length 783;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLDWGR 6  
 |:||||  
 Db 540 CIDWGR 545

#### RESULT 4

##### ACP1\_SPIOL

ID ACP1\_SPIOL STANDARD; PRT; 138 AA.  
 AC P07854;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Acyl carrier protein I, chloroplast precursor (ACP I).  
 GN ACL1.1.  
 OS Spinacia oleracea (Spinach).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OX NCBI\_TaxID=3562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Scherer D.E., Knauf V.C.;  
 RT "Isolation of a cDNA clone for the acyl carrier protein-I of  
 RT spinach."  
 RL Plant Mol. Biol. 9:127-134(1987).  
 RN [2]  
 RP SEQUENCE OF 57-138.  
 RC TISSUE=Leaf;  
 RX MEDLINE=85021451; PubMed=6486822;  
 RA Kuo T.M., Ohlrogge J.B.;  
 RT "The primary structure of spinach acyl carrier protein."  
 RL Arch. Biochem. Biophys. 234:290-296(1984).  
 CC -!- FUNCTION: Carrier of the growing fatty acid chain in fatty acid  
 CC biosynthesis.  
 CC -!- PATHWAY: De novo fatty acid biosynthesis.  
 CC -!- PTM: 4'-phosphopantetheine is transferred from CoA to a specific  
 CC serine of apo-ACP by acpS. This modification is essential for  
 CC activity because fatty acids are bound in thioester linkage to the  
 CC sulfhydryl of the prosthetic group (By similarity).  
 CC -!- SIMILARITY: Contains 1 acyl carrier domain.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17636; AAA34023.1; -.
DR PIR; A28052; AYSP.
DR HSSP; P02901; IACP.
DR InterPro; IPR003231; Acyl_carrier.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00550; pp-binding; 1.
DR ProDom; PD000887; Acyl_carrier; 1.
DR TIGRFAMs; TIGR00517; acyl_carrier; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Phosphopantetheine; Chloroplast;
KW Transit peptide; Multigene family.
FT TRANSIT      1      56      CHLOROPLAST.
FT CHAIN        57     138     ACYL CARRIER PROTEIN I.
FT BINDING      94      94     PHOSPHOPANTETHEINE.
FT CONFLICT     66      66     C -> S (IN REF. 2).
SQ SEQUENCE    138 AA;  14909 MW;  B3FB8F08BF657980 CRC64;

```

```

Query Match          70.4%;  Score 38;  DB 1;  Length 138;
Best Local Similarity 83.3%;  Pred. No. 7.8;
Matches      5;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

QY      1 CLDWGR 6
        |||||:
Db      34 CLDWGK 39

```

```

RESULT 5
EAR_ASFB7
ID EAR_ASFB7      STANDARD;      PRT;   179 AA.
AC P42485;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apoptosis regulator Bcl-2 homolog precursor.
GN A179L.
OS African swine fever virus (strain BA71V) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10498;
RN [1]
RP SEQUENCE FROM N.A.
RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA Rodriguez J.F., Vinuela E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
RT virus.";
RL Virology 208:249-278(1995).
CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.

```



CC    -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
CC       CYCLE.  
CC    -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
CC    -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
CC    -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
CC       VIRUS BHRF1.  
CC    -----  
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC    the European Bioinformatics Institute. There are no restrictions on its  
CC    use by non-profit institutions as long as its content is in no way  
CC    modified and this statement is not removed. Usage by and for commercial  
CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC    -----  
DR    EMBL; U18466; AAA65271.1; -.  
DR    InterPro; IPR000712; Bcl2\_BH.  
DR    InterPro; IPR002475; BCL2\_family.  
DR    Pfam; PF00452; Bcl-2; 1.  
DR    SMART; SM00337; BCL; 1.  
DR    PROSITE; PS50062; BCL2\_FAMILY; 1.  
DR    PROSITE; PS01080; BH1; 1.  
DR    PROSITE; PS01258; BH2; 1.  
KW    Signal; Apoptosis.  
FT    SIGNAL           1       18           POTENTIAL.  
FT    CHAIN           19       179          APOPTOSIS REGULATOR BCL-2 HOMOLOG.  
FT    DOMAIN           76       95          BH1.  
FT    DOMAIN           126      141         BH2.  
SQ    SEQUENCE    179 AA;   21075 MW;   62CB13D82374BF35 CRC64;

Query Match                   70.4%;   Score 38;   DB 1;   Length 179;  
Best Local Similarity   71.4%;   Pred. No. 9.8;  
Matches       5;   Conservative   2;   Mismatches   0;   Indels    0;   Gaps       0;

Qy           2 LDWGRIC 8  
              ::|||||  
Db           82 INWGRIC 88

# RESULT 6

## EAR\_ASFE4

ID   EAR\_ASFE4           STANDARD;       PRT;    179 AA.  
AC   Q07818;  
DT   01-FEB-1995 (Rel. 31, Created)  
DT   01-FEB-1995 (Rel. 31, Last sequence update)  
DT   28-FEB-2003 (Rel. 41, Last annotation update)  
DE   Apoptosis regulator Bcl-2 homolog precursor (LMH-5W).  
GN   LMW5-HL.  
OS   African swine fever virus (strain E-70 / isolate MS44) (ASFV).  
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
OX   NCBI\_TaxID=39014;  
RN   [1]  
RP   SEQUENCE FROM N.A.  
RX   MEDLINE=93287262; PubMed=8389936;  
RA   Neilan J.G., Lu Z., Afonso C.L., Kutish G.F., Sussman M.D., Rock D.L.;  
RT   "An African swine fever virus gene with similarity to the  
RT   proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1.";

RL J. Virol. 67:4391-4394(1993).  
 CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
 CC CYCLE.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
 CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
 CC VIRUS BHRF1.  
 DR InterPro; IPR000712; Bcl2\_BH.  
 DR InterPro; IPR002475; BCL2\_family.  
 DR Pfam; PF00452; Bcl-2; 1.  
 DR SMART; SM00337; BCL; 1.  
 DR PROSITE; PS50062; BCL2\_FAMILY; 1.  
 DR PROSITE; PS01080; BH1; 1.  
 DR PROSITE; PS01258; BH2; 1.  
 KW Signal; Apoptosis.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 179 APOPTOSIS REGULATOR BCL-2 HOMOLOG.  
 FT DOMAIN 76 95 BH1.  
 FT DOMAIN 126 141 BH2.  
 SQ SEQUENCE 179 AA; 21131 MW; 56B1C22790677BD2 CRC64;

Query Match 70.4%; Score 38; DB 1; Length 179;  
 Best Local Similarity 71.4%; Pred. No. 9.8;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 :|||  
 Db 82 INWGRIC 88

#### RESULT 7

EAR\_ASFM2  
 ID EAR\_ASFM2 STANDARD; PRT; 179 AA.  
 AC Q07819;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Apoptosis regulator Bcl-2 homolog precursor (LMH-5W).  
 GN LMW5-HL.  
 OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).  
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
 OX NCBI\_TaxID=10500;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93287262; PubMed=8389936;  
 RA Neilan J.G., Lu Z., Afonso C.L., Kutish G.F., Sussman M.D., Rock D.L.;  
 RT "An African swine fever virus gene with similarity to the  
 RT proto-oncogene bcl-2 and the Epstein-Barr virus gene BHRF1."  
 RL J. Virol. 67:4391-4394(1993).  
 CC -!- FUNCTION: SUPPRESSION OF APOPTOSIS IN HOST CELLS.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY AND LATE IN THE INFECTION  
 CC CYCLE.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
 CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY. SIMILAR TO EPSTEIN-BARR  
 CC VIRUS BHRF1.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09548; AAA17034.1; -.
DR InterPro; IPR000712; Bcl2_BH.
DR InterPro; IPR002475; BCL2_family.
DR Pfam; PF00452; Bcl-2; 1.
DR SMART; SM00337; BCL; 1.
DR PROSITE; PS50062; BCL2_FAMILY; 1.
DR PROSITE; PS01080; BH1; 1.
DR PROSITE; PS01258; BH2; 1.
KW Signal; Apoptosis.
FT SIGNAL          1      18      POTENTIAL.
FT CHAIN           19     179      APOPTOSIS REGULATOR BCL-2 HOMOLOG.
FT DOMAIN          76      95      BH1.
FT DOMAIN         126     141      BH2.
SQ SEQUENCE       179 AA;  21068 MW;  0A4204D5643C66E4 CRC64;

```

```

Query Match          70.4%;  Score 38;  DB 1;  Length 179;
Best Local Similarity 71.4%;  Pred. No. 9.8;
Matches      5;  Conservative    2;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      2 LDWGRIC 8
        ::|||||
Db      82 INWGRIC 88

```

# RESULT 8

## ARCC\_RHIET

```

ID ARCC_RHIET      STANDARD;      PRT;   310 AA.
AC O31019;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamate kinase (EC 2.7.2.2).
GN ARCC.
OS Rhizobium etli.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=29449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053854; PubMed=9393705;
RA D'Hooghe I., Vander Wauwe C., Michiels J., Tricot C., de Wilde P.,
RA Vanderleyden J., Stalon V.;
RT "The arginine deiminase pathway in Rhizobium etli: DNA sequence
RT analysis and functional study of the arcABC genes.";
RL J. Bacteriol. 179:7403-7409(1997).
CC -!- CATALYTIC ACTIVITY: ATP + NH(3) + CO(2) = ADP + carbamoyl
CC phosphate.
CC -!- PATHWAY: Arginine degradation via arginine deiminase; third step.

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: Belongs to the carbamate kinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF025543; AAC46020.1; -.  
 DR HSSP; P95474; 1E19.  
 DR InterPro; IPR001048; Aa\_kinase.  
 DR InterPro; IPR003964; Bac\_carb\_kinase.  
 DR Pfam; PF00696; aakinase; 1.  
 DR PRINTS; PR01469; CARBMTKINASE.  
 DR TIGRFAMs; TIGR00746; arcC; 1.  
 KW Transferase; Kinase; Arginine metabolism.  
 SQ SEQUENCE 310 AA; 33504 MW; 50115ABC1D597224 CRC64;

Query Match 70.4%; Score 38; DB 1; Length 310;  
 Best Local Similarity 83.3%; Pred. No. 16;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
 |||||:  
 Db 233 CLDWGK 238

# RESULT 9

## KMOS\_RAT

ID KMOS\_RAT STANDARD; PRT; 339 AA.  
 AC P00539;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Proto-oncogene serine/threonine-protein kinase mos (EC 2.7.1.37)  
 DE (c-mos).  
 GN MOS.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84144095; PubMed=6322135;  
 RA van der Hoorn F.A., Firzlauff J.;  
 RT "Complete c-mos (rat) nucleotide sequence: presence of conserved  
 RT domains in c-mos proteins."  
 RL Nucleic Acids Res. 12:2147-2156(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Skeletal muscle;  
 RX MEDLINE=90363547; PubMed=1697408;  
 RA Leibovitch S.A., Lenormand J.-L., Leibovitch M.-P., Guiller M.,  
 RA Mallard L., Harel J.;

RT "Rat myogenic c-mos cDNA: cloning sequence analysis and regulation  
 RT during muscle development.";  
 RL Oncogene 5:1149-1157(1990).  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- TISSUE SPECIFICITY: MOS IS EXPRESSED MAINLY IN GONADAL TISSUES,  
 CC AND CARDIAC AND SKELETAL MUSCLES.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X00422; CAA25123.1; -.  
 DR EMBL; X52952; CAA37128.1; -.  
 DR PIR; A00648; TVRTM.  
 DR HSSP; P08631; 1AD5.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; Proto-oncogene;  
 KW ATP-binding.  
 FT DOMAIN 61 335 PROTEIN KINASE.  
 FT NP\_BIND 67 75 ATP (BY SIMILARITY).  
 FT BINDING 88 88 ATP (BY SIMILARITY).  
 FT ACT\_SITE 196 196 BY SIMILARITY.  
 FT CONFLICT 47 47 L -> V (IN REF. 2).  
 FT CONFLICT 102 102 R -> A (IN REF. 2).  
 SQ SEQUENCE 339 AA; 37621 MW; A074246A5E471278 CRC64;

Query Match 70.4%; Score 38; DB 1; Length 339;  
 Best Local Similarity 57.1%; Pred. No. 18;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LDWGRIC 8  
 :|||::|  
 Db 56 IDWGQVC 62

RESULT 10  
 CDL5\_HUMAN  
 ID CDL5\_HUMAN STANDARD; PRT; 418 AA.  
 AC Q14004;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cell division cycle 2-like protein kinase 5 (EC 2.7.1.-)  
 DE (Cholinesterase-related cell division controller) (CDC2-related  
 DE protein kinase 5).  
 GN CDC2L5 OR CDC2L OR CHED.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Glioblastoma;  
 RX MEDLINE=92115704; PubMed=1731328;  
 RA Lapidot-Lifson Y., Patinkin D., Prody C.A., Ehrlich G., Seidman S.,  
 RA Ben-Aziz R., Benseler F., Eckstein F., Zakut H., Soreq H.;  
 RT "Cloning and antisense oligodeoxynucleotide inhibition of a human  
 RT homolog of cdc2 required in hematopoiesis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:579-583(1992).  
 CC -!- FUNCTION: MAY BE A CONTROLLER OF THE MITOTIC CELL CYCLE. INVOLVED  
 CC IN THE BLOOD CELL DEVELOPMENT.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN, LIVER, MUSCLE AND IN  
 CC ADULT BRAIN. ALSO EXPRESSED IN NEUROBLASTOMA AND GLIOBLASTOMA  
 CC TUMORS.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CDC2/CDKX SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M80629; AAA58424.1; -.  
 DR HSSP; P24941; 1BUH.  
 DR Genew; HGNC:1733; CDC2L5.  
 DR GK; Q14004; -.  
 DR MIM; 603309; -.  
 DR GO; GO:0007275; P:development; TAS.  
 DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.  
 DR GO; GO:0007088; P:regulation of mitosis; TAS.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 91 384 PROTEIN KINASE.  
 FT NP\_BIND 97 105 ATP (BY SIMILARITY).  
 FT BINDING 120 120 ATP (BY SIMILARITY).  
 FT ACT\_SITE 223 223 BY SIMILARITY.  
 SQ SEQUENCE 418 AA; 48211 MW; 4EBA77F1C48CD915 CRC64;

Query Match 70.4%; Score 38; DB 1; Length 418;  
 Best Local Similarity 57.1%; Pred. No. 21;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
:| | | : : |  
Db 81 IDWGKLC 87

RESULT 11

RS5\_METTH

ID RS5\_METTH STANDARD; PRT; 216 AA.  
AC O26131;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S5P.  
GN RPS5P OR MTH23.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics."  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -!- FUNCTION: With S4 and S12 plays an important role in translational  
CC accuracy (By similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S4  
CC (By similarity).  
CC -!- DOMAIN: The N-terminal domain interacts with the head of the 30S  
CC subunit; the C-terminal domain interacts with the body and  
CC contacts protein S4. The interaction surface between S4 and S5 is  
CC involved in control of translational fidelity.  
CC -!- SIMILARITY: Contains 1 S5 DRBM domain.  
CC -!- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000796; AAB84532.1; -.  
DR PIR; E69128; E69128.  
DR HSSP; P02357; 1PKP.  
DR HAMAP; MF\_01307; -; 1.  
DR InterPro; IPR000851; Ribosomal\_S5.

DR InterPro; IPR005324; Ribosomal\_S5\_C.  
 DR InterPro; IPR005711; S5\_euk\_arch.  
 DR Pfam; PF00333; Ribosomal\_S5; 1.  
 DR Pfam; PF03719; Ribosomal\_S5\_C; 1.  
 DR TIGRFAMs; TIGR01020; rpsE\_arch; 1.  
 DR PROSITE; PS00585; RIBOSOMAL\_S5; 1.  
 DR PROSITE; PS50881; S5\_DSRBD; 1.  
 KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
 FT DOMAIN 51 114 S5 DRBM.  
 SQ SEQUENCE 216 AA; 23626 MW; FC9E7D051BBB7565 CRC64;

Query Match 68.5%; Score 37; DB 1; Length 216;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 .| ||| :|  
 Db 118 CGDWGCVC 125

# RESULT 12

## AM3A\_ORYSA

ID AM3A\_ORYSA STANDARD; PRT; 440 AA.  
 AC P27932;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-amylase isozyme 3A precursor (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).  
 GN AMY1.2 OR AMY3A.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Japonica M202; TISSUE=Etiolated leaf;  
 RX MEDLINE=91329692; PubMed=1714318;  
 RA Sutliff T.D., Huang N., Litts J.C., Rodriguez R.L.;  
 RT "Characterization of an alpha-amylase multigene cluster in rice.";  
 RL Plant Mol. Biol. 16:579-591(1991).  
 CC -!- FUNCTION: IMPORTANT FOR BREAKDOWN OF ENDOSPERM STARCH DURING GERMINATION.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.  
 CC -!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN EMBRYO-DERIVED CALLUS TISSUE.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT A HIGH LEVEL DURING GERMINATION IN THE ALEURONES CELLS UNDER THE CONTROL OF THE PLANT HORMONE GIBBERELIC ACID AND IN THE DEVELOPING GRAINS AT A LOW LEVEL.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X56336; CAA39776.1; -.  
DR PIR; S14958; S14958.  
DR HSSP; P04063; 1AVA.  
DR Gramene; P27932; -.  
DR InterPro; IPR006589; Alp\_amyl\_cat\_sub.  
DR InterPro; IPR006047; Alpha\_amyl\_cat.  
DR InterPro; IPR006046; Glyco\_hydro\_13.  
DR Pfam; PF00128; alpha-amylase; 1.  
DR PRINTS; PR00110; ALPHAAMYLASE.  
DR SMART; SM00642; Aamy; 1.  
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal;  
KW Multigene family.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 440 ALPHA-AMYLASE ISOZYME 3A.  
FT ACT\_SITE 207 207 BY SIMILARITY.  
FT ACT\_SITE 315 315 BY SIMILARITY.  
FT METAL 119 119 CALCIUM (BY SIMILARITY).  
FT METAL 178 178 CALCIUM (BY SIMILARITY).  
SQ SEQUENCE 440 AA; 48872 MW; 5E9B78C29AA91C2B CRC64;

Query Match 68.5%; Score 37; DB 1; Length 440;  
Best Local Similarity 70.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy 1 CLDWG--RIC 8  
||||| ||  
Db 143 CLDWGPSMIC 152

#### RESULT 13

##### GATD\_ARCFU

ID GATD\_ARCFU STANDARD; PRT; 418 AA.  
AC O29380;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamyl-tRNA(Gln) amidotransferase subunit D (EC 6.3.5.-) (Glu-ADT  
DE subunit D).  
GN GATD OR AF0882.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 CC -!- FUNCTION: Allows the formation of correctly charged Gln-tRNA(Gln)  
 CC through the transamidation of misacylated Glu-tRNA(Gln) in  
 CC organisms which lack glutaminyl-tRNA synthetase. The reaction  
 CC takes place in the presence of glutamine and ATP through an  
 CC activated gamma-phospho-Glu-tRNA(Gln). The gatDE system is  
 CC specific for glutamate and does not act on aspartate (By  
 CC similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
 CC + phosphate + L-glutaminyl-tRNA(Gln) + L-glutamate.  
 CC -!- SUBUNIT: Heterodimer of gatD and gatE (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ASPARAGINASE 1 FAMILY. GATD SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001043; AAB90360.1; -.  
 DR PIR; B69360; B69360.  
 DR HSSP; P00805; 3ECA.  
 DR TIGR; AF0882; -.  
 DR HAMAP; MF\_00586; -; 1.  
 DR InterPro; IPR006033; AsnASEI.  
 DR InterPro; IPR006034; Asp/Glutamnse.  
 DR Pfam; PF00710; Asparaginase; 1.  
 DR PRINTS; PR00139; ASNGLNASE.  
 DR ProDom; PD003221; Asp/Glutamnse; 1.  
 DR TIGRFAMs; TIGR00519; asnASE\_I; 1.  
 DR PROSITE; PS00144; ASN\_GLN\_ASE\_1; 1.  
 DR PROSITE; PS00917; ASN\_GLN\_ASE\_2; 1.  
 KW Protein biosynthesis; Ligase; Complete proteome.  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT ACT\_SITE 166 166 BY SIMILARITY.  
 FT ACT\_SITE 167 167 BY SIMILARITY.  
 FT ACT\_SITE 243 243 BY SIMILARITY.  
 SQ SEQUENCE 418 AA; 47091 MW; C3F4A4AE831DD05D CRC64;

Query Match 67.6%; Score 36.5; DB 1; Length 418;  
 Best Local Similarity 75.0%; Pred. No. 38;  
 Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CLDWGRIC 8  
 || |||:|  
 Db 352 CL-WGRVC 358

RESULT 14

DCHS\_ENTAE

ID DCHS\_ENTAE STANDARD; PRT; 377 AA.

AC P28577;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Histidine decarboxylase (EC 4.1.1.22) (HDC).

GN HDC.

OS Enterobacter aerogenes (Aerobacter aerogenes).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

OX NCBI\_TaxID=548;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91236707; PubMed=2033044;

RA Kamath A.V., Vaaler G.L., Snell E.E.;

RT "Pyridoxal phosphate-dependent histidine decarboxylases. Cloning,  
RT sequencing, and expression of genes from Klebsiella planticola and  
RT Enterobacter aerogenes and properties of the overexpressed enzymes.";

RL J. Biol. Chem. 266:9432-9437(1991).

CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).

CC -!- COFACTOR: Pyridoxal phosphate.

CC -!- SUBUNIT: Homotetramer (By similarity).

CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,  
CC GAD, HDC AND TYRDC).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M62745; AAA24802.1; -.

DR PIR; A40004; A40004.

DR HAMAP; MF\_00609; -; 1.

DR InterPro; IPR002129; Pyridoxal\_deC.

DR Pfam; PF00282; pyridoxal\_deC; 1.

DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.

KW Lyase; Decarboxylase; Pyridoxal phosphate.

FT INIT\_MET 0 0 BY SIMILARITY.

FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).

SQ SEQUENCE 377 AA; 42303 MW; 4C7A3334ACA7D6AE CRC64;

Query Match 66.7%; Score 36; DB 1; Length 377;

Best Local Similarity 62.5%; Pred. No. 42;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

| | | | |

Db 49 CGDWGEYC 56

RESULT 15

DCHS\_KLEPL

ID DCHS\_KLEPL STANDARD; PRT; 377 AA.  
AC P28578; Q8KHD1; Q8KHF6;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Histidine decarboxylase (EC 4.1.1.22) (HDC).  
GN HDC.  
OS Klebsiella planticola (Raoultella planticola).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Raoultella.  
OX NCBI\_TaxID=575;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 43176;  
RX MEDLINE=91236707; PubMed=2033044;  
RA Kamath A.V., Vaaler G.L., Snell E.E.;  
RT "Pyridoxal phosphate-dependent histidine decarboxylases. Cloning,  
RT sequencing, and expression of genes from Klebsiella planticola and  
RT Enterobacter aerogenes and properties of the overexpressed enzymes.";  
RL J. Biol. Chem. 266:9432-9437(1991).  
RN [2]  
RP SEQUENCE OF 90-317 FROM N.A.  
RC STRAIN=19-3, 27-1, 28-1, 42-1, S8, and Y1-1;  
RX MEDLINE=22083483; PubMed=12089029;  
RA Kanki M., Yoda T., Tsukamoto T., Shibata T.;  
RT "Klebsiella pneumoniae produces no histamine: Raoultella planticola  
RT and Raoultella ornithinolytica strains are histamine producers.";  
RL Appl. Environ. Microbiol. 68:3462-3466(2002).  
CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).  
CC -!- COFACTOR: Pyridoxal phosphate.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- MISCELLANEOUS: This histamine-producing bacteria (HPB) causes  
CC histamine fish poisoning.  
CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,  
CC GAD, HDC AND TYRDC).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M62746; AAA25071.1; -.  
DR EMBL; AB075216; BAB97305.1; -.  
DR EMBL; AB075217; BAB97306.1; -.  
DR EMBL; AB075218; BAB97307.1; -.  
DR EMBL; AB075219; BAB97308.1; -.  
DR EMBL; AB075220; BAB97309.1; -.  
DR EMBL; AB075221; BAB97310.1; -.  
DR PIR; B40004; B40004.  
DR HAMAP; MF\_00609; -; 1.  
DR InterPro; IPR002129; Pyridoxal\_deC.  
DR Pfam; PF00282; pyridoxal\_deC; 1.

DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
KW Lyase; Decarboxylase; Pyridoxal phosphate.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).  
FT VARIANT 147 147 A -> T (IN STRAINS 28-1 AND 42-1).  
FT VARIANT 183 183 Q -> E (IN STRAINS 28-1 AND 42-1).  
FT CONFLICT 155 155 R -> A (IN REF. 1).  
SQ SEQUENCE 377 AA; 42766 MW; 131A20A0A540D25A CRC64;

Query Match 66.7%; Score 36; DB 1; Length 377;  
Best Local Similarity 62.5%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| | | | |  
Db 49 CGDWGEYC 56

Search completed: November 13, 2003, 09:46:37  
Job time : 5.58333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 21.0833 Seconds  
(without alignments)  
97.917 Million cell updates/sec

Title: US-09-228-866-8  
Perfect score: 54  
Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*

8: sp\_organelle:\*  
 9: sp\_phage:\*  
 10: sp\_plant:\*  
 11: sp\_rodent:\*  
 12: sp\_virus:\*  
 13: sp\_vertebrate:\*  
 14: sp\_unclassified:\*  
 15: sp\_rvirus:\*  
 16: sp\_bacteriap:\*  
 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	45	83.3	62	4	Q9P1A5	Q9pla5 homo sapien
2	41	75.9	134	9	Q38422	Q38422 bacterioph
3	41	75.9	247	12	Q8V223	Q8v223 potato leaf
4	41	75.9	247	12	Q8QHN2	Q8qhn2 potato leaf
5	41	75.9	247	12	Q8V236	Q8v236 potato leaf
6	41	75.9	247	12	Q8V227	Q8v227 potato leaf
7	41	75.9	247	12	Q8V234	Q8v234 potato leaf
8	41	75.9	247	12	Q8V238	Q8v238 potato leaf
9	41	75.9	247	12	Q8QYQ9	Q8qyq9 potato leaf
10	41	75.9	247	12	Q8QYP0	Q8qyp0 potato leaf
11	41	75.9	247	12	Q8V225	Q8v225 potato leaf
12	41	75.9	247	12	Q8V229	Q8v229 potato leaf
13	41	75.9	247	12	Q8V244	Q8v244 potato leaf
14	41	75.9	247	12	Q8UYD3	Q8uyd3 potato leaf
15	41	75.9	247	12	Q8QYP7	Q8qyp7 potato leaf
16	41	75.9	247	12	Q8V242	Q8v242 potato leaf
17	41	75.9	247	12	Q8V231	Q8v231 potato leaf
18	41	75.9	247	12	Q8QYR7	Q8qyr7 potato leaf
19	41	75.9	247	12	Q8V240	Q8v240 potato leaf
20	41	75.9	247	12	Q84835	Q84835 potato leaf
21	41	75.9	269	12	Q8QYN2	Q8qyn2 potato leaf
22	40	74.1	518	10	Q94HA3	Q94ha3 oryza sativ
23	39	72.2	213	13	O57503	O57503 scelopor
24	39	72.2	690	5	Q9XWC5	Q9xwc5 caenorhabdi
25	38	70.4	1452	4	Q9H4A0	Q9h4a0 homo sapien
26	38	70.4	1512	4	Q9H4A1	Q9h4a1 homo sapien
27	37	68.5	64	16	Q9A5U9	Q9a5u9 caulobacter
28	37	68.5	72	5	Q8MVM6	Q8mvm6 boltenia vi
29	37	68.5	72	12	Q8VB81	Q8vb81 white spot
30	37	68.5	216	13	Q9PUP1	Q9pup1 guira guira
31	37	68.5	414	16	Q97MZ7	Q97mz7 clostridium
32	37	68.5	631	5	Q9W271	Q9w271 drosophila
33	37	68.5	698	5	Q961X2	Q961x2 drosophila
34	37	68.5	727	5	Q9W270	Q9w270 drosophila
35	37	68.5	789	5	Q9W269	Q9w269 drosophila
36	37	68.5	810	10	Q8LN78	Q8ln78 oryza sativ
37	37	68.5	990	16	Q8EAY1	Q8eay1 shewanella

38	37	68.5	1092	3	Q9UVY2	Q9uvy2 pneumocysti
39	36	66.7	135	16	Q8EJF3	Q8ejf3 shewanella
40	36	66.7	170	16	Q9RIS7	Q9ris7 streptomyce
41	36	66.7	233	16	Q8KBN0	Q8kbn0 chlorobium
42	36	66.7	256	12	Q993H3	Q993h3 callitrichi
43	36	66.7	261	17	Q8U291	Q8u291 pyrococcus
44	36	66.7	271	10	Q9FMY4	Q9fmy4 arabidopsis
45	36	66.7	275	13	O13090	O13090 pleurodeles

# ALIGNMENTS

## RESULT 1

Q9P1A5

ID Q9P1A5 PRELIMINARY; PRT; 62 AA.  
AC Q9P1A5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE PRO0889.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
RA He F.;  
RT "Functional prediction of the coding sequences of 79 new genes deduced  
RT by analysis of cDNA clones from human fetal liver.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF119839; AAF69593.1; -.  
SQ SEQUENCE 62 AA; 6643 MW; 478EE1DC006A36E7 CRC64;

Query Match 83.3%; Score 45; DB 4; Length 62;  
Best Local Similarity 100.0%; Pred. No. 0.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	2 LDWGRIC 8
Db	28 LDWGRIC 34

## RESULT 2

Q38422

ID Q38422 PRELIMINARY; PRT; 134 AA.  
AC Q38422;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE ORF 1.  
OS Bacteriophage SP01.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC SP01-like viruses.  
OX NCBI\_TaxID=10685;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92267370; PubMed=1587473;  
 RA Scarlato V., Sayre M.H.;  
 RT "Sequence of the bacteriophage SP01 gene 30.";  
 RL Gene 114:115-119(1992).  
 DR EMBL; M82842; AAA32596.1; -.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR PROSITE; PS00334; MYB\_2; 1.  
 SQ SEQUENCE 134 AA; 15251 MW; C7F3911875DFCD7E CRC64;

Query Match 75.9%; Score 41; DB 9; Length 134;  
 Best Local Similarity 71.4%; Pred. No. 9.5;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 ||||:|  
 Db 115 LDWGKVC 121

#### RESULT 3

Q8V223

ID Q8V223 PRELIMINARY; PRT; 247 AA.  
 AC Q8V223;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Potato leaf roll virus isolate Br1 P0.  
 OS Potato leafroll virus (PLrV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Br1;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Analysis of the genetic diversity of potato leafroll virus reveals  
 RT major evolutionary events and differential selection pressures between  
 RT overlapping reading frame products.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453406; AAL73431.1; -.  
 SQ SEQUENCE 247 AA; 28155 MW; 04A4FAB4430BB95B CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:| :|  
 Db 84 CLEWGLLC 91

#### RESULT 4

Q8QHN2

ID Q8QHN2 PRELIMINARY; PRT; 247 AA.  
 AC Q8QHN2;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)



DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Strain Zim13, complete genome (Strain Fr1, complete genome).  
 OS Potato leafroll virus (PLrV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Zim13, and Fr1;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Sequence analysis of Potato leafroll virus isolates reveals genetic  
 RT stability, major evolutionary events and differential selection  
 RT pressure between overlapping reading frame products."  
 RL J. Gen. Virol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Zim13, and Fr1;  
 RA Guyader S., Giblot Ducray D.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453388; AAL77913.1; -.  
 DR EMBL; AF453391; AAL77937.1; -.  
 SQ SEQUENCE 247 AA; 28193 MW; ED8770198ED5A657 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:|| :|  
 Db 84 CLEWGLLC 91

# RESULT 5

Q8V236

ID Q8V236 PRELIMINARY; PRT; 247 AA.  
 AC Q8V236;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Potato leaf roll virus isolate K5 P0.  
 OS Potato leafroll virus (PLrV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K5;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Analysis of the genetic diversity of potato leafroll virus reveals  
 RT major evolutionary events and differential selection pressures between  
 RT overlapping reading frame products."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453399; AAL73417.1; -.  
 SQ SEQUENCE 247 AA; 28137 MW; B5AB2FB1D8BDFBEA CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;

Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

RESULT 6

Q8V227

ID Q8V227 PRELIMINARY; PRT; 247 AA.  
AC Q8V227;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate L13D P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L13D;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals  
RT major evolutionary events and differential selection pressures between  
RT overlapping reading frame products.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453404; AAL73427.1; -.  
SQ SEQUENCE 247 AA; 28285 MW; 5E29D6B5CB3CA550 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

RESULT 7

Q8V234

ID Q8V234 PRELIMINARY; PRT; 247 AA.  
AC Q8V234;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate L18 P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L18;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals

RT major evolutionary events and differential selection pressures between  
RT overlapping reading frame products.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453400; AAL73419.1; -.  
SQ SEQUENCE 247 AA; 28240 MW; 0F906E24D27C6E42 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||:||:|  
Db 84 CLEWGLLC 91

RESULT 8

Q8V238

ID Q8V238 PRELIMINARY; PRT; 247 AA.  
AC Q8V238;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate Au252 P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Au252;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals  
RT major evolutionary events and differential selection pressures between  
RT overlapping reading frame products.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453398; AAL73415.1; -.  
SQ SEQUENCE 247 AA; 28066 MW; D2D316E7B8A2CBA0 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||:||:|  
Db 84 CLEWGLLC 91

RESULT 9

Q8QYQ9

ID Q8QYQ9 PRELIMINARY; PRT; 247 AA.  
AC Q8QYQ9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Strain Noir, complete genome.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Noir;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Sequence analysis of Potato leafroll virus isolates reveals genetic  
 RT stability, major evolutionary events and differential selection  
 RT pressure between overlapping reading frame products.";  
 RL J. Gen. Virol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Noir;  
 RA Guyader S., Giblot Ducray D.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453390; AAL77929.1; -.  
 SQ SEQUENCE 247 AA; 27998 MW; 2CA32449F308E958 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:|| :|  
 Db 84 CLEWGLLC 91

# RESULT 10

## Q8QYP0

ID Q8QYP0 PRELIMINARY; PRT; 247 AA.  
 AC Q8QYP0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Strain CU87, complete genome.  
 OS Potato leafroll virus (PLrV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CU87;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Sequence analysis of Potato leafroll virus isolates reveals genetic  
 RT stability, major evolutionary events and differential selection  
 RT pressure between overlapping reading frame products.";  
 RL J. Gen. Virol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CU87;  
 RA Guyader S., Giblot Ducray D.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453393; AAL77952.1; -.  
 SQ SEQUENCE 247 AA; 28065 MW; 0A7FC9B41625A664 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

RESULT 11

Q8V225

ID Q8V225 PRELIMINARY; PRT; 247 AA.  
AC Q8V225;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate 14.1 P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=14.1;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals  
RT major evolutionary events and differential selection pressures between  
RT overlapping reading frame products."  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453405; AAL73429.1; -.  
SQ SEQUENCE 247 AA; 28126 MW; 77E5A0EEA3241924 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

RESULT 12

Q8V229

ID Q8V229 PRELIMINARY; PRT; 247 AA.  
AC Q8V229;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate L13B P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L13B;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals  
RT major evolutionary events and differential selection pressures between

RT overlapping reading frame products.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453403; AAL73425.1; -.  
SQ SEQUENCE 247 AA; 28182 MW; B04F27626A31C250 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

#### RESULT 13

Q8V244

ID Q8V244 PRELIMINARY; PRT; 247 AA.  
AC Q8V244;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate Au16 P0.  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Polerovirus.  
OX NCBI\_TaxID=12045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Au16;  
RA Guyader S., Giblot Ducray D.;  
RT "Analysis of the genetic diversity of potato leafroll virus reveals  
RT major evolutionary events and differential selection pressures between  
RT overlapping reading frame products.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF453395; AAL73409.1; -.  
SQ SEQUENCE 247 AA; 28094 MW; 3C5F0F7A96D94245 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

#### RESULT 14

Q8UYD3

ID Q8UYD3 PRELIMINARY; PRT; 247 AA.  
AC Q8UYD3;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Potato leaf roll virus isolate 1457 P0 (Potato leaf roll virus isolate  
DE L7 P0).  
OS Potato leafroll virus (PLrV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1457, and L7;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Analysis of the genetic diversity of potato leafroll virus reveals  
 RT major evolutionary events and differential selection pressures between  
 RT overlapping reading frame products.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF454283; AAL73433.1; -.  
 DR EMBL; AF453401; AAL73421.1; -.  
 SQ SEQUENCE 247 AA; 28164 MW; 96653CA43E190587 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:||:|  
 Db 84 CLEWGLLC 91

#### RESULT 15

Q8QYP7

ID Q8QYP7 PRELIMINARY; PRT; 247 AA.  
 AC Q8QYP7;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Strain CIP01, complete genome.  
 OS Potato leafroll virus (PLrV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CIP01;  
 RA Guyader S., Giblot Ducray D.;  
 RT "Sequence analysis of Potato leafroll virus isolates reveals genetic  
 RT stability, major evolutionary events and differential selection  
 RT pressure between overlapping reading frame products.";  
 RL J. Gen. Virol. 0:0-0(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CIP01;  
 RA Guyader S., Giblot Ducray D.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF453392; AAL77945.1; -.  
 SQ SEQUENCE 247 AA; 28142 MW; B5E0277D75260591 CRC64;

Query Match 75.9%; Score 41; DB 12; Length 247;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 ||:||:|

Db            84 CLEWGLLC 91

Search completed: November 13, 2003, 09:51:08  
Job time : 23.0833 secs



OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 9.5 Seconds  
 (without alignments)  
 35.630 Million cell updates/sec

Title: US-09-228-866-8  
 Perfect score: 54  
 Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	54	100.0	8	1	US-08-526-710-8	Sequence 8, Appli
2	54	100.0	8	3	US-08-862-855-8	Sequence 8, Appli
3	54	100.0	8	3	US-09-226-985-8	Sequence 8, Appli
4	54	100.0	8	4	US-09-227-906-8	Sequence 8, Appli
5	48	88.9	8	1	US-08-526-710-7	Sequence 7, Appli
6	48	88.9	8	3	US-08-862-855-7	Sequence 7, Appli
7	48	88.9	8	3	US-09-226-985-7	Sequence 7, Appli
8	48	88.9	8	4	US-09-227-906-7	Sequence 7, Appli
9	38	70.4	10	2	US-08-733-505A-35	Sequence 35, Appl
10	38	70.4	10	2	US-08-706-741B-70	Sequence 70, Appl
11	38	70.4	10	2	US-08-924-695A-70	Sequence 70, Appl

12	38	70.4	20	1	US-08-248-819A-39	Sequence 39, Appl
13	38	70.4	20	2	US-08-337-646A-57	Sequence 57, Appl
14	38	70.4	20	3	US-08-927-326-57	Sequence 57, Appl
15	38	70.4	21	1	US-08-112-208C-15	Sequence 15, Appl
16	38	70.4	21	1	US-08-248-819A-17	Sequence 17, Appl
17	38	70.4	21	2	US-08-337-646A-35	Sequence 35, Appl
18	38	70.4	21	2	US-08-856-531-15	Sequence 15, Appl
19	38	70.4	21	2	US-08-856-034-15	Sequence 15, Appl
20	38	70.4	21	3	US-08-927-326-35	Sequence 35, Appl
21	38	70.4	21	4	US-09-379-820A-15	Sequence 15, Appl
22	38	70.4	67	1	US-08-321-071A-11	Sequence 11, Appl
23	38	70.4	179	1	US-08-607-269-27	Sequence 27, Appl
24	38	70.4	179	5	PCT-US95-04600-27	Sequence 27, Appl
25	38	70.4	187	1	US-08-471-058-17	Sequence 17, Appl
26	38	70.4	187	3	US-08-471-057-17	Sequence 17, Appl
27	38	70.4	187	4	US-08-470-865-17	Sequence 17, Appl
28	36	66.7	139	3	US-08-930-894-6	Sequence 6, Appli
29	35	64.8	228	4	US-09-107-532A-4878	Sequence 4878, Ap
30	34.5	63.9	662	4	US-09-198-452A-169	Sequence 169, App
31	34	63.0	157	4	US-09-996-243-103	Sequence 103, App
32	34	63.0	337	4	US-09-252-991A-22646	Sequence 22646, A
33	34	63.0	436	2	US-08-576-626A-47	Sequence 47, Appl
34	34	63.0	462	3	US-09-036-987A-18	Sequence 18, Appl
35	34	63.0	462	3	US-09-370-700-18	Sequence 18, Appl
36	34	63.0	462	4	US-09-603-207-18	Sequence 18, Appl
37	34	63.0	3567	2	US-07-642-734C-4	Sequence 4, Appli
38	34	63.0	3567	3	US-08-439-009A-4	Sequence 4, Appli
39	33	61.1	50	4	US-09-461-325-152	Sequence 152, App
40	33	61.1	211	1	US-08-321-071A-16	Sequence 16, Appl
41	33	61.1	288	4	US-09-252-991A-29594	Sequence 29594, A
42	33	61.1	398	4	US-09-252-991A-24881	Sequence 24881, A
43	33	61.1	508	3	US-09-111-730-2	Sequence 2, Appli
44	33	61.1	631	4	US-09-252-991A-26007	Sequence 26007, A
45	33	61.1	871	4	US-09-328-352-7076	Sequence 7076, Ap

#### ALIGNMENTS

RESULT 1  
 US-08-526-710-8  
 ; Sequence 8, Application US/08526710  
 ; Patent No. 5622699  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruoslahti, Erkki  
 ; APPLICANT: Pasqualini, Renata  
 ; TITLE OF INVENTION: Method of Identifying Molecules That  
 ; TITLE OF INVENTION: Home to a Selected Organ In Vivo  
 ; NUMBER OF SEQUENCES: 44  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell and Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 92122  
 ; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-8

```

```

Query Match          100.0%; Score 54; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CLDWGRIC 8
        |||||
Db      1 CLDWGRIC 8

```

# RESULT 2

```

US-08-862-855-8
; Sequence 8, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-8

```

```

Query Match          100.0%; Score 54; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 CLDWGRIC 8
        |||||
Db      1 CLDWGRIC 8

```

# RESULT 3

US-09-226-985-8

```

; Sequence 8, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-8

```

```

Query Match          100.0%; Score 54; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLDWGRIC 8
        |||||
Db      1 CLDWGRIC 8

```

#### RESULT 4

US-09-227-906-8

```

; Sequence 8, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-8

```

```

Query Match          100.0%; Score 54; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CLDWGRIC 8
        |||||
Db      1 CLDWGRIC 8

```

# RESULT 5

US-08-526-710-7

```

; Sequence 7, Application US/08526710
; Patent No. 5622699
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-7

```

```

Query Match          88.9%; Score 48; DB 1; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLDWGRIC 8
        | |||||
Db      1 CKDWGRIC 8

```

# RESULT 6

US-08-862-855-7

```

; Sequence 7, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-7

```

```

Query Match          88.9%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLDWGRIC 8
        | |||||
Db      1 CKDWGRIC 8

```

# RESULT 7

US-09-226-985-7

```

; Sequence 7, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997

```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-7

```

```

Query Match          88.9%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLDWGRIC 8
        | |||||
Db      1 CKDWGRIC 8

```

```

RESULT 8
US-09-227-906-7
; Sequence 7, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-7

```

```

Query Match          88.9%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLDWGRIC 8
        | |||||
Db      1 CKDWGRIC 8

```

# RESULT 9

US-08-733-505A-35

; Sequence 35, Application US/08733505A

; Patent No. 5856445

; GENERAL INFORMATION:

; APPLICANT: KORSMEYER, STANLEY J.

; TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF

; TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HOWELL & HAFERKAMP, L.C.

; STREET: 7733 FORSYTH BLVD., SUITE 1400

; CITY: ST. LOUIS

; STATE: MISSOURI

; COUNTRY: USA

; ZIP: 63105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/733,505A

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: HOLLAND, DONALD R.

; REGISTRATION NUMBER: 35,197

```

; REFERENCE/DOCKET NUMBER: 965458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-733-505A-35

```

```

Query Match          70.4%; Score 38; DB 2; Length 10;
Best Local Similarity 71.4%; Pred. No. 1.2;
Matches      5; Conservative      2; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 LDWGRIC 8
        ::|||||
Db      3 INWGRIC 9

```

RESULT 10

US-08-706-741B-70

```

; Sequence 70, Application US/08706741B
; Patent No. 5955593
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, STANLEY J.
; TITLE OF INVENTION: BH3 INTERACTING DOMAIN DEATH AGONIST
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63146
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,741B
; FILING DATE: 09-SEP-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 965017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid

```

; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-706-741B-70

Query Match 70.4%; Score 38; DB 2; Length 10;  
Best Local Similarity 71.4%; Pred. No. 1.2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
::|||||  
Db 3 INWGRIC 9

RESULT 11

US-08-924-695A-70

; Sequence 70, Application US/08924695A  
; Patent No. 5998583  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, STANLEY J.  
; TITLE OF INVENTION: BH3 INTERACTING DOMAIN DEATH AGONIST  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MISSOURI  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/924,695A  
; FILING DATE: 09-SEP-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 971798  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314) 727-5188  
; TELEFAX: (314) 727-6092  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-924-695A-70

Query Match 70.4%; Score 38; DB 2; Length 10;  
Best Local Similarity 71.4%; Pred. No. 1.2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
:|||||  
Db 3 INWGRIC 9

RESULT 12

US-08-248-819A-39

; Sequence 39, Application US/08248819A  
; Patent No. 5700638  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, Stanley J.  
; TITLE OF INVENTION: CELL DEATH REGULATORS  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Khourie and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/248,819A  
; FILING DATE: 25-NAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/112,208  
; FILING DATE: 26-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 15726A-000610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 4  
; OTHER INFORMATION: /note= "Amino acid is either K  
; OTHER INFORMATION:  
US-08-248-819A-39

Query Match 70.4%; Score 38; DB 1; Length 20;  
Best Local Similarity 71.4%; Pred. No. 2.4;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
:|||||  
Db 7 INWGRIC 13

RESULT 13

US-08-337-646A-57

; Sequence 57, Application US/08337646A  
; Patent No. 5856171  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, Stanley J.  
; TITLE OF INVENTION: CELL DEATH REGULATORS  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/337,646A  
; FILING DATE: 10-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/248,819  
; FILING DATE: 25-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/112,208  
; FILING DATE: 26-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 15726A-000620  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 4  
; OTHER INFORMATION: /note= "Amino acid is either K  
; OTHER INFORMATION:

US-08-337-646A-57

Query Match 70.4%; Score 38; DB 2; Length 20;  
Best Local Similarity 71.4%; Pred. No. 2.4;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
::|||||  
Db 7 INWGRIC 13

RESULT 14

US-08-927-326-57

; Sequence 57, Application US/08927326

; Patent No. 6184202

; GENERAL INFORMATION:

; APPLICANT: KORSMEYER, Stanley J.

; TITLE OF INVENTION: CELL DEATH REGULATORS

; NUMBER OF SEQUENCES: 78

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend Khourie and Crew

; STREET: 379 Lytton Avenue

; CITY: Palo Alto

; STATE: California

; COUNTRY: US

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,326

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/337,646

; FILING DATE: 10-NOV-1994

; APPLICATION NUMBER: US 08/248,819

; FILING DATE: 25-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/112,208

; FILING DATE: 26-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Smith, William M

; REGISTRATION NUMBER: 30,223

; REFERENCE/DOCKET NUMBER: 15726A-000620

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 326-2400

; TELEFAX: (415) 326-2422

; INFORMATION FOR SEQ ID NO: 57:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Region

; LOCATION: 4  
; OTHER INFORMATION: /note= "Amino acid is either K  
; OTHER INFORMATION:  
US-08-927-326-57

Query Match 70.4%; Score 38; DB 3; Length 20;  
Best Local Similarity 71.4%; Pred. No. 2.4;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
::|||||  
Db 7 INWGRIC 13

RESULT 15

US-08-112-208C-15

; Sequence 15, Application US/08112208C  
; Patent No. 5691179  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, Stanley J.  
; TITLE OF INVENTION: CELL DEATH REGULATORS  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Khourie and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/112,208C  
; FILING DATE: 26-AUG-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 15726A-000610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 5  
; OTHER INFORMATION: /note= "Amino acid is either K  
; OTHER INFORMATION:



US-08-112-208C-15

Query Match 70.4%; Score 38; DB 1; Length 21;  
Best Local Similarity 71.4%; Pred. No. 2.5;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
::|||||  
Db 8 INWGRIC 14

Search completed: November 13, 2003, 09:54:58  
Job time : 9.5 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 9.5 Seconds  
(without alignments)  
35.630 Million cell updates/sec

Title: US-09-228-866-9  
Perfect score: 46  
Sequence: 1 CTRITESC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	46	100.0	8	1	US-08-526-710-9	Sequence 9, Appli	
2	46	100.0	8	3	US-08-862-855-9	Sequence 9, Appli	
3	46	100.0	8	3	US-09-226-985-9	Sequence 9, Appli	
4	46	100.0	8	4	US-09-227-906-9	Sequence 9, Appli	
5	33	71.7	90	4	US-09-860-793-5	Sequence 5, Appli	
6	33	71.7	501	4	US-09-157-257-8	Sequence 8, Appli	
7	33	71.7	539	4	US-09-157-257-6	Sequence 6, Appli	
8	32	69.6	48	4	US-09-240-078-1	Sequence 1, Appli	
9	32	69.6	50	3	US-09-031-902-2	Sequence 2, Appli	
10	32	69.6	54	1	US-08-757-541-8	Sequence 8, Appli	
11	32	69.6	54	3	US-09-033-275-8	Sequence 8, Appli	

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-8

```

```

Query Match          87.3%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CKDWGRIC 8
        | |||||
Db      1 CLDWGRIC 8

```

```

RESULT 9
US-08-454-196-11
; Sequence 11, Application US/08454196
; Patent No. 5770361
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUTKA-MALEN, SYLVIE
; APPLICANT: EVERS, STEFAN
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: PROTEIN CONFERRING AN INDUCIBLE
; TITLE OF INVENTION: RESISTANCE TO GLYCOPEPTIDES, PARTICULARLY IN GRAM-
POSITIVE
; TITLE OF INVENTION: BACTERIA
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/454,196
; FILING DATE: 07-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/15671
; FILING DATE: 18-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93/08356
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-101-O PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-454-196-11

```

```

Query Match          70.9%; Score 39; DB 1; Length 306;
Best Local Similarity 85.7%; Pred. No. 27;
Matches      6; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

```

```

Qy      1 CKDWGRI 7
        || ||||
Db      250 CKGWGRI 256

```

# RESULT 10

US-08-286-819A-33

; Sequence 33, Application US/08286819A

; Patent No. 5871910

; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL

; APPLICANT: DUKTA-MALEN, SYLVIE

; APPLICANT: MOLINAS, CATHERINE

; APPLICANT: COURVALIN, PATRICE

; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN

PARTICULAR

; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING  
FOR

; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 S. Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

```

; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,819A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5871910man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-819A-33

```

```

Query Match          70.9%;  Score 39;  DB 2;  Length 306;
Best Local Similarity 85.7%;  Pred. No. 27;
Matches      6;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy      1 CKDWGRI 7
        || ||||
Db      250 CKGWGRI 256

```

```

RESULT 11
US-08-980-357-33
; Sequence 33, Application US/08980357
; Patent No. 6013508
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE

```

; APPLICANT: MOLINAS, CATHERINE  
 ; APPLICANT: COURVALIN, PATRICE  
 ; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
 ; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN  
 PARTICULAR  
 ; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING  
 FOR  
 ; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
 ; NUMBER OF SEQUENCES: 54  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/980,357  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/286,819  
 ; FILING DATE: 05-AUG-1994  
 ; APPLICATION NUMBER: US 08/174,682  
 ; FILING DATE: 28-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/917,146  
 ; FILING DATE: 10-AUG-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/FR/91/00855  
 ; FILING DATE: 29-OCT-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: FR 9013579  
 ; FILING DATE: 31-OCT-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 6013508man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 413-3000  
 ; TELEFAX: (703) 413-2220  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 33:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 306 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-980-357-33

Query Match

70.9%; Score 39; DB 3; Length 306;

Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 250 CKGWGRI 256

RESULT 12

US-09-064-033-11

; Sequence 11, Application US/09064033

; Patent No. 6087106

; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL

; APPLICANT: DUTKA-MALEN, SYLVIE

; APPLICANT: EVERS, STEFAN

; APPLICANT: COURVALIN, PATRICE

; TITLE OF INVENTION: PROTEIN CONFERRING AN INDUCIBLE

; TITLE OF INVENTION: RESISTANCE TO GLYCOPEPTIDES, PARTICULARLY IN GRAM-  
POSITIVE

; TITLE OF INVENTION: BACTERIA

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,033

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/454,196

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 93/08356

; FILING DATE: 07-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 660-101-O PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 306 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-064-033-11

Query Match 70.9%; Score 39; DB 3; Length 306;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CKDWGRI 7  
|| ||||  
Db 250 CKGWGRI 256

RESULT 13

US-09-291-046-11

; Sequence 11, Application US/09291046  
; Patent No. 6569622

; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL  
; DUTKA-MALEN, SYLVIE  
; EVERS, STEFAN  
; COURVALIN, PATRICE

; TITLE OF INVENTION: PROTEIN CONFERRING AN INDUCIBLE  
; RESISTANCE TO GLYCOPEPTIDES, PARTICULARLY IN GRAM-  
POSITIVE

; BACTERIA

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; P.C.

; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/291,046

; FILING DATE: 14-Apr-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/454,196

; FILING DATE: <Unknown>

; APPLICATION NUMBER: FR 93/08356

; FILING DATE: 07-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 660-101-O PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 11:



```

;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 306 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-291-046-11

```

```

Query Match          70.9%;  Score 39;  DB 4;  Length 306;
Best Local Similarity 85.7%;  Pred. No. 27;
Matches      6;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy      1 CKDWGRI 7
        |||||
Db      250 CKGWGRI 256

```

```

RESULT 14
US-09-328-352-7076
; Sequence 7076, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7076
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7076

```

```

Query Match          67.3%;  Score 37;  DB 4;  Length 871;
Best Local Similarity 50.0%;  Pred. No. 1.7e+02;
Matches      4;  Conservative    3;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy      1 CKDWGRIC 8
        |:|:|
Db      319 CRDWFQLC 326

```

```

RESULT 15
US-08-733-505A-35
; Sequence 35, Application US/08733505A
; Patent No. 5856445
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, STANLEY J.
; TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF
; TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.

```

```

; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,505A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 965458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-733-505A-35

```

```

Query Match          65.5%;  Score 36;  DB 2;  Length 10;
Best Local Similarity 83.3%;  Pred. No. 2.6;
Matches      5;  Conservative    1;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      3 DWGRIC 8
        :|||||
Db      4 NWGRIC 9

```

```

Search completed: November 13, 2003, 09:54:58
Job time : 10.5 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 26.9167 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-8  
Perfect score: 54  
Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	54	100.0	8	18	AAW13419		Brain homing pepti
2	54	100.0	8	21	AAB07394		Brain homing pepti
3	54	100.0	8	22	AAE11800		Phage peptide #8 t
4	54	100.0	8	23	AAU10711		Brain homing pepti
5	48	88.9	8	18	AAW13418		Brain homing pepti
6	48	88.9	8	21	AAB07393		Brain homing pepti
7	48	88.9	8	22	AAE11799		Phage peptide #7 t
8	48	88.9	8	23	AAU10710		Brain homing pepti
9	42	77.8	704	22	ABG19897		Novel human diagno
10	41	75.9	121	22	AAU48487		Propionibacterium
11	40	74.1	208	22	ABG00095		Novel human diagno
12	39	72.2	77	22	AAO13632		Human polypeptide
13	39	72.2	80	22	AAU16925		Human novel secret
14	38	70.4	20	22	AAB74174		LMW5-HL BH1 domain
15	38	70.4	21	20	AAW87835		Bcl-2 related prot
16	38	70.4	21	22	AAB74152		LMW5-HL BH1 domain
17	38	70.4	52	23	ABP35141		Human ORF4114 prot
18	38	70.4	137	9	AAP81138		Sequence of plant
19	38	70.4	137	15	AAR54978		Spinach acyl carri
20	38	70.4	138	22	ABB03987		Human musculoskele
21	38	70.4	138	24	ABU13281		Novel human muscul
22	38	70.4	296	24	ABP57005		Thiobacillus ferro
23	38	70.4	331	23	ABB77060		Human protein sequ
24	37	68.5	91	23	ABP58934		Human focal adhesi
25	37	68.5	727	22	ABB59835		Drosophila melanog
26	37	68.5	789	22	ABB59802		Drosophila melanog
27	37	68.5	849	22	ABB59837		Drosophila melanog
28	36	66.7	30	22	AAU05849		Cone snail O-supe
29	36	66.7	77	22	AAU05867		Cone snail O-supe
30	36	66.7	94	22	AAU42306		Propionibacterium
31	36	66.7	108	22	AAM84755		Human immune/haema
32	36	66.7	109	22	ABG21437		Novel human diagno
33	36	66.7	173	21	AAB38201		Human secreted pro
34	36	66.7	430	22	ABB68578		Drosophila melanog
35	36	66.7	435	20	AAZ29954		Human CG1CE short
36	36	66.7	450	21	AAG34061		Zea mays protein f
37	36	66.7	453	22	ABB65852		Drosophila melanog
38	36	66.7	462	22	ABB64544		Drosophila melanog
39	36	66.7	485	21	AAG34060		Zea mays protein f
40	36	66.7	510	21	AAG34059		Zea mays protein f
41	36	66.7	531	23	AAU75055		Arabidopsis short-
42	35	64.8	26	22	AAU05850		Cone snail O-supe
43	35	64.8	26	22	AAU05868		Cone snail O-supe
44	35	64.8	62	23	ABP04123		Human ORFX protein
45	35	64.8	64	22	ABG52965		Human liver peptid

# ALIGNMENTS

RESULT 1

AAW13419

ID AAW13419 standard; Peptide; 8 AA.

XX

AC AAW13419;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 15; Page 68; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
CC peptide that was identified using a novel method for obtaining  
CC molecules that home to a selected organ or tissue. This in vivo  
CC panning method typically involves administering a phage display  
CC library to a subject, and identifying expressed peptides which  
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
CC vascular tissue or tumour tissue. The isolated peptides (see  
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
CC labels to the selected organ/tissue (claimed) or to identify and/or  
CC isolate target molecules (claimed). The peptides can be directly  
CC identified in vivo, as compared to prior art in vitro screening  
CC methods, which require further examination to see if they maintain  
CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 54; DB 18; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

Db

|||||||  
1 CLDWGRIC 8

RESULT 2

AAB07394

ID AAB07394 standard; peptide; 8 AA.

XX

AC AAB07394;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 8.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..8

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in  
PT vivo panning comprises administering a library of diverse peptides  
PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was  
CC identified by using in vivo panning to screen a library of potential  
CC organ homing molecules. The present sequence can be used to direct a  
CC moiety to a the brain tissue, by linking the moiety to the present  
CC sequence. Examples of potential moieties are drugs, toxins or a  
CC detectable label. The present sequence contains a DXXR amino acid motif  
CC (AAB12027). The DXXR motif resembles the RGD, DGR and NGR motifs that  
CC bind to certain integrins.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 54; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
|||  
Db 1 CLDWGRIC 8

RESULT 3

AAE11800

ID AAE11800 standard; peptide; 8 AA.

XX

AC AAE11800;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #8 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo

PT panning that selectively home to a selected organ or tissue useful for

PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 54; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
|||  
Db 1 CLDWGRIC 8

RESULT 4

AAU10711

ID AAU10711 standard; peptide; 8 AA.

XX

AC AAU10711;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #8 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;

KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer



CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 54; DB 23; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLDWGRIC 8  
||| |||  
Db 1 CLDWGRIC 8

# RESULT 5

AAW13418

ID AAW13418 standard; Peptide; 8 AA.

XX

AC AAW13418;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (IJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 15; Page 68; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing

CC peptide that was identified using a novel method for obtaining  
 CC molecules that home to a selected organ or tissue. This in vivo  
 CC panning method typically involves administering a phage display  
 CC library to a subject, and identifying expressed peptides which  
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
 CC vascular tissue or tumour tissue. The isolated peptides (see  
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
 CC labels to the selected organ/tissue (claimed) or to identify and/or  
 CC isolate target molecules (claimed). The peptides can be directly  
 CC identified in vivo, as compared to prior art in vitro screening  
 CC methods, which require further examination to see if they maintain  
 CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 88.9%; Score 48; DB 18; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
 | | | | |  
 Db 1 CKDWGRIC 8

# RESULT 6

AAB07393

ID AAB07393 standard; peptide; 8 AA.

XX

AC AAB07393;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 7.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..8

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in  
PT vivo panning comprises administering a library of diverse peptides  
PT linked to a tag which facilitates recovery of these peptides -  
XX  
PS Example 2; Column 17; 20pp; English.  
XX  
CC The present sequence is a mouse brain homing peptide. This sequence was  
CC identified by using in vivo panning to screen a library of potential  
CC organ homing molecules. The present sequence can be used to direct a  
CC moiety to a the brain tissue, by linking the moiety to the present  
CC sequence. Examples of potential moieties are drugs, toxins or a  
CC detectable label. The present sequence contains a DXXR amino acid motif  
CC (AAB12027). The DXXR motif resembles the RGD, DGR and NGR motifs that  
CC bind to certain integrins.  
XX  
SQ Sequence 8 AA;

Query Match 88.9%; Score 48; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| | | | | | |  
Db 1 CKDWGRIC 8

#### RESULT 7

AAE11799

ID AAE11799 standard; peptide; 8 AA.

XX

AC AAE11799;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #7 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;  
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo  
PT panning that selectively home to a selected organ or tissue useful for  
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 8 AA;

Query Match 88.9%; Score 48; DB 22; Length 8;

Best Local Similarity 87.5%; Pred. No. 9.3e+05;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

| | | | |

Db 1 CKDWGRIC 8

RESULT 8

AAU10710

ID AAU10710 standard; peptide; 8 AA.

XX

AC AAU10710;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #7 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 88.9%; Score 48; DB 23; Length 8;  
Best Local Similarity 87.5%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

| | | | | | |

Db 1 CKDWGRIC 8

#### RESULT 9

ABG19897

ID ABG19897 standard; Protein; 704 AA.

XX

AC ABG19897;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #19888.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS84084.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 50256; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 704 AA;

Query Match 77.8%; Score 42; DB 22; Length 704;  
 Best Local Similarity 85.7%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 ||||:||  
 Db 35 LDWGKIC 41

RESULT 10  
 AAU48487  
 ID AAU48487 standard; Protein; 121 AA.

XX  
 AC AAU48487;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #9383.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59542.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 9682; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 121 AA;

Query Match 75.9%; Score 41; DB 22; Length 121;

Best Local Similarity 62.5%; Pred. No. 22;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

|::| |||

Db 92 CVEWSRIC 99

#### RESULT 11

ABG00095

ID ABG00095 standard; Protein; 208 AA.

XX

AC ABG00095;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #86.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS64282.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX

PS Claim 20; SEQ ID No 30454; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques



CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 208 AA;

Query Match 74.1%; Score 40; DB 22; Length 208;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
|:|||||  
Db 90 LNWGRIC 96

# RESULT 12

AAO13632

ID AAO13632 standard; Protein; 77 AA.

XX

AC AAO13632;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 27524.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US04927.

XX

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-514838/56.  
 DR N-PSDB; AAI93563.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 20; SEQ ID NO 27524; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 77 AA;

Query Match 72.2%; Score 39; DB 22; Length 77;  
 Best Local Similarity 85.7%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LDWGRIC 8  
 | | | | |  
 Db 5 LSWGRIC 11

# RESULT 13

AAU16925

ID AAU16925 standard; Protein; 80 AA.

XX

AC AAU16925;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human novel secreted protein, SEQ ID 166.

XX

KW Human; immunosuppressive; antiarthritic; antirheumatic;  
 KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnerary; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.

XX

OS Homo sapiens.

XX

PN WO200155441-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01320.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-476222/51.

DR N-PSDB; AAS26830.

XX

PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
PT diagnose diseases or disorders associated with aberrant expression or  
PT activity of polypeptides, for treating blood clotting disorder,  
PT haemophilia -

XX

PS Claim 11; SEQ ID No 166; 601pp; English.

XX

CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present

Query Match 72.2%; Score 39; DB 22; Length 80;  
 Best Local Similarity 71.4%; Pred. No. 32;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRI 7  
 ||||| :  
 Db 54 CLDWGHV 60

#### RESULT 14

AAB74174

ID AAB74174 standard; Peptide; 20 AA.

XX

AC AAB74174;

XX

DT 22-MAY-2001 (first entry)

XX

DE LMW5-HL BH1 domain #2.

XX

KW Bax; cytostatic; immunosuppressive; immunostimulant; infection;  
 KW apoptosis modulator; bcl-2 associated X protein; cancer therapy; AIDS;  
 KW autoimmunity; immunodeficiency; reperfusion injury; stroke; aging;  
 KW myocardial infarction; traumatic brain injury; ischaemia; Bcl-2;  
 KW neurodegenerative diseases; hepatitis; transplant rejection; toxemia;  
 KW lymphoproliferative disease.

XX

OS Unidentified.

XX

PN US6184202-B1.

XX

PD 06-FEB-2001.

XX

PF 11-SEP-1997; 97US-0927326.

XX

PR 10-NOV-1994; 94US-0337646.

PR 26-AUG-1993; 93US-0112208.

PR 25-MAY-1994; 94US-0248819.

XX

PA (UNIW ) UNIV WASHINGTON.

XX

PI Korsmeyer SJ;

XX

DR WPI; 2001-256104/26.

XX

PT Modulating apoptosis of a cell, useful in maintaining homeostasis in  
 PT adult tissues, or treating proliferative or autoimmune diseases,  
 PT comprises administering a bcl-2 polypeptide that interacts with a 21 kD

PT bcl-2 associated X protein -  
 XX  
 PS Example 11; Fig 22; 105pp; English.  
 XX  
 CC The present invention relates to a method of modulating apoptosis of a  
 CC cell. The method comprises administering to the cell an agent,  
 CC comprising a BH1 domain or BH2 domain, capable of modulating formation of  
 CC at least one complex selected from bcl-2:bcl-2 complexes, bcl-XL:bcl-XL  
 CC complexes, bcl-2 associated X protein (Bax):Bax complexes, bcl-2:Bax  
 CC complexes or bcl-XL:Bax complexes. Modulating apoptosis is especially  
 CC useful in cancer therapy, and treating autoimmunity, immunodeficiency  
 CC diseases (e.g. AIDS), reperfusion injury, myocardial infarction, stroke,  
 CC traumatic brain injury, neurodegenerative diseases, aging, ischaemia,  
 CC toxemia, infection, hepatitis, transplant rejection, and  
 CC lymphoproliferative diseases. The present sequence is a peptide, which  
 CC was used in the method of the present invention.  
 XX  
 SQ Sequence 20 AA;

Query Match 70.4%; Score 38; DB 22; Length 20;  
 Best Local Similarity 71.4%; Pred. NO. 12;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 : : | | | |  
 Db 7 INWGRIC 13

# RESULT 15

AAW87835

ID AAW87835 standard; Peptide; 21 AA.

XX

AC AAW87835;

XX

DT 10-MAR-1999 (first entry)

XX

DE Bcl-2 related protein (LMW5-HL) domain BH1 peptide.

XX

KW Bcl-2 related protein; Bax; bcl-2; modulator; domain BH1;

KW bcl-2-related function; apoptosis; dimer; Bcl-xL; Mcl-1; A1.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Misc-difference 5

FT /note= "Arg or Lys"

XX

PN US5856171-A.

XX

PD 05-JAN-1999.

XX

PF 10-NOV-1994; 94US-0337646.

XX

PR 10-NOV-1994; 94US-0337646.

PR 26-AUG-1993; 93US-0112208.

PR 25-MAY-1994; 94US-0248819.

XX

PA (UNIW ) UNIV WASHINGTON.  
 XX  
 PI Korsmeyer SJ;  
 XX  
 DR WPI; 1999-105119/09.  
 XX  
 PT DNA composition encoding bcl-2 two-hybrid and reporter system - for  
 PT identifying modulators of bcl-2 function  
 XX  
 PS Example 10; Fig 14A; 105pp; English.  
 XX  
 CC AAW87832-36 represent the amino acid sequences of domain BH1 of  
 CC Bcl-2-related proteins. The specification describes a composition  
 CC comprising a hybrid protein comprising an activator domain of a  
 CC transcriptional activator protein and a bcl-2 family member having  
 CC a BH1 domain and a BH2 domain; another hybrid protein comprising a  
 CC DNA-binding domain of the transcriptional activator protein and a  
 CC second bcl-2 family member having a BH1 domain and a BH2 domain; and  
 CC a reporter gene linked to a transcriptional regulatory element whose  
 CC transcriptional activity is dependent on the presence or absence of  
 CC a dimer of the two hybrid proteins. The bcl-2 family members are  
 CC selected from naturally occurring Bcl-2, Bcl-xL, Bax, Mcl-1, A1,  
 CC fragments thereof, and mutants having a mutation in the BH1 and/or  
 CC BH2 domain that alters intermolecular binding of the two bcl-2 family  
 CC members. The composition is used to identify modulators of bcl-2-related  
 CC function, e.g. substances that inhibit binding of Bax to bcl-2, which  
 CC would be potentially useful as drugs for modulating apoptosis.  
 XX  
 SQ Sequence 21 AA;  
  
 Query Match 70.4%; Score 38; DB 20; Length 21;  
 Best Local Similarity 71.4%; Pred. No. 13;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 :|||  
 Db 8 INWGRIC 14

Search completed: November 13, 2003, 09:45:27  
 Job time : 26.9167 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 16.5833 Seconds  
 (without alignments)  
 88.069 Million cell updates/sec

Title: US-09-228-866-8  
 Perfect score: 54  
 Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62



Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	39	72.2	80	9	US-09-764-898-166	Sequence 166, App
2	38	70.4	21	15	US-10-277-693A-15	Sequence 15, Appl
3	38	70.4	138	10	US-09-764-877-1934	Sequence 1934, Ap
4	38	70.4	187	15	US-10-101-482-17	Sequence 17, Appl
5	38	70.4	296	15	US-10-186-886-11	Sequence 11, Appl
6	38	70.4	418	10	US-09-771-161A-218	Sequence 218, App
7	36	66.7	30	10	US-09-749-637A-92	Sequence 92, Appl
8	36	66.7	77	10	US-09-749-637A-119	Sequence 119, App
9	36	66.7	219	15	US-10-156-761-11025	Sequence 11025, A
10	35	64.8	26	10	US-09-749-637A-93	Sequence 93, Appl
11	35	64.8	26	10	US-09-749-637A-120	Sequence 120, App
12	35	64.8	64	9	US-09-864-761-43385	Sequence 43385, A
13	35	64.8	146	9	US-09-925-299-996	Sequence 996, App
14	35	64.8	146	11	US-09-925-299-996	Sequence 996, App
15	35	64.8	381	10	US-09-738-626-5815	Sequence 5815, Ap

16	35	64.8	807	9	US-09-847-046-2	Sequence 2, Appli
17	35	64.8	1091	12	US-10-032-585-7431	Sequence 7431, Ap
18	35	64.8	1509	10	US-09-901-940-2	Sequence 2, Appli
19	34	63.0	29	10	US-09-749-637A-95	Sequence 95, Appl
20	34	63.0	29	10	US-09-749-637A-104	Sequence 104, App
21	34	63.0	30	10	US-09-749-637A-89	Sequence 89, Appl
22	34	63.0	45	10	US-09-911-150-5	Sequence 5, Appli
23	34	63.0	53	12	US-10-029-386-33402	Sequence 33402, A
24	34	63.0	61	11	US-09-764-891-2990	Sequence 2990, Ap
25	34	63.0	75	10	US-09-749-637A-51	Sequence 51, Appl
26	34	63.0	76	9	US-09-814-122-66	Sequence 66, Appl
27	34	63.0	133	15	US-10-156-761-15084	Sequence 15084, A
28	34	63.0	157	9	US-09-989-722-103	Sequence 103, App
29	34	63.0	157	9	US-09-989-723-103	Sequence 103, App
30	34	63.0	157	9	US-09-989-279-103	Sequence 103, App
31	34	63.0	157	9	US-09-989-727-103	Sequence 103, App
32	34	63.0	157	10	US-09-989-731-103	Sequence 103, App
33	34	63.0	157	10	US-09-989-732-103	Sequence 103, App
34	34	63.0	157	10	US-09-991-073-103	Sequence 103, App
35	34	63.0	157	10	US-09-990-442-103	Sequence 103, App
36	34	63.0	157	10	US-09-991-163-103	Sequence 103, App
37	34	63.0	157	10	US-09-993-604-103	Sequence 103, App
38	34	63.0	157	10	US-09-990-456-103	Sequence 103, App
39	34	63.0	157	10	US-09-989-721-103	Sequence 103, App
40	34	63.0	157	10	US-09-992-598-103	Sequence 103, App
41	34	63.0	157	10	US-09-989-293A-103	Sequence 103, App
42	34	63.0	157	10	US-09-989-735-103	Sequence 103, App
43	34	63.0	157	10	US-09-990-444-103	Sequence 103, App
44	34	63.0	157	10	US-09-991-181-103	Sequence 103, App
45	34	63.0	157	10	US-09-989-730-103	Sequence 103, App

#### ALIGNMENTS

##### RESULT 1

US-09-764-898-166

```
; Sequence 166, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ01
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 166
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-166
```

```
Query Match          72.2%; Score 39; DB 9; Length 80;
Best Local Similarity 71.4%; Pred. No. 22;
Matches      5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 CLDWGRI 7  
| | | | :  
Db 54 CLDWGHV 60

RESULT 2

US-10-277-693A-15

; Sequence 15, Application US/10277693A  
; Publication No. US20030096367A1  
; GENERAL INFORMATION:  
; APPLICANT: Korsmeyer, Stanley J.  
; TITLE OF INVENTION: Cell Death Agonists  
; FILE REFERENCE: 56029/36280  
; CURRENT APPLICATION NUMBER: US/10/277,693A  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/379,820  
; PRIOR FILING DATE: 1999-08-24  
; PRIOR APPLICATION NUMBER: 08/112,208  
; PRIOR FILING DATE: 1993-08-26  
; PRIOR APPLICATION NUMBER: 08/856,034  
; PRIOR FILING DATE: 1997-05-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Murine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (5)..(5)  
; OTHER INFORMATION:  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (5)..(5)  
; OTHER INFORMATION: Amino acid is either K (Lys) or R (Arg)  
US-10-277-693A-15

Query Match 70.4%; Score 38; DB 15; Length 21;  
Best Local Similarity 71.4%; Pred. No. 10;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
: | | | |  
Db 8 INWGRIC 14

RESULT 3

US-09-764-877-1934

; Sequence 1934, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1934  
; LENGTH: 138  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-877-1934

Query Match 70.4%; Score 38; DB 10; Length 138;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRI 7  
|| ||||  
Db 125 CLGWGRI 131

RESULT 4

US-10-101-482-17

; Sequence 17, Application US/10101482

; Publication No. US20030008837A1

; GENERAL INFORMATION:

; APPLICANT: KIEFER, MICHAEL C.

; BARR, PHILIP J.

; TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA

; ENCODING THE PROTEINS AND METHODS OF USE THEREOF

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/101,482

; FILING DATE: 18-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/320,157

; FILING DATE: 07-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: LEHNHARDT, SUSAN K.

; REGISTRATION NUMBER: 33,943

; REFERENCE/DOCKET NUMBER: 23647-20007.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

```
;          LENGTH: 187 amino acids
;          TYPE: amino acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;          SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-101-482-17
```

```
Query Match          70.4%;  Score 38;  DB 15;  Length 187;
Best Local Similarity 71.4%;  Pred. No. 70;
Matches      5;  Conservative      2;  Mismatches      0;  Indels      0;  Gaps      0;
```

```
Qy          2 LDWGRIC 8
             ::|||||
Db          95 INWGRIC 101
```

RESULT 5

US-10-186-886-11

```
; Sequence 11, Application US/10186886
; Publication No. US20030119061A1
; GENERAL INFORMATION:
; APPLICANT: Navia, Manuel A.
; APPLICANT: Ala, Paul J.
; APPLICANT: Griffith, James P.
; APPLICANT: Ali, Janid A.
; APPLICANT: Faerman, Carlos H.
; APPLICANT: Moe, Scott T.
; APPLICANT: Magee, Andrew S.
; APPLICANT: Connelly, Patrick R.
; APPLICANT: Perola, Emanuele
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS
ANTIBACTERIAL
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 10283-014001
; CURRENT APPLICATION NUMBER: US/10/186,886
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/301,676
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Thiobacillus ferrooxidans
US-10-186-886-11
```

```
Query Match          70.4%;  Score 38;  DB 15;  Length 296;
Best Local Similarity 71.4%;  Pred. No. 1e+02;
Matches      5;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;
```

```
Qy          1 CLDWGRI 7
             | ||||:
Db          241 CRDWGRV 247
```

RESULT 6

US-09-771-161A-218  
 ; Sequence 218, Application US/09771161A  
 ; Patent No. US20020110811A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEVINE, et al.  
 ; TITLE OF INVENTION: VARIANTS Of PROTEIN KINASES  
 ; FILE REFERENCE: 802620-2005.1  
 ; CURRENT APPLICATION NUMBER: US/09/771,161A  
 ; CURRENT FILING DATE: 2001-01-26  
 ; PRIOR APPLICATION NUMBER: 09/724,676  
 ; PRIOR FILING DATE: 2000-11-28  
 ; PRIOR APPLICATION NUMBER: 136776  
 ; PRIOR FILING DATE: 2000-06-15  
 ; PRIOR APPLICATION NUMBER: 135619  
 ; PRIOR FILING DATE: 2000-04-12  
 ; NUMBER OF SEQ ID NOS: 273  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 218  
 ; LENGTH: 418  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-771-161A-218

Query Match 70.4%; Score 38; DB 10; Length 418;  
 Best Local Similarity 57.1%; Pred. No. 1.4e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
 :|||::|  
 Db 81 IDWGKLC 87

RESULT 7  
 US-09-749-637A-92  
 ; Sequence 92, Application US/09749637A  
 ; Patent No. US20020173449A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: University of Utah Research Foundation  
 ; APPLICANT: Cognetix, Inc.  
 ; APPLICANT: Olivera, Baldomero M.  
 ; APPLICANT: Cartier, G. Edward  
 ; APPLICANT: Watkins, Maren  
 ; APPLICANT: Hillyard, David R.  
 ; APPLICANT: McIntosh, J. Michael  
 ; APPLICANT: Layer, Richard T.  
 ; APPLICANT: Jones, Robert M.  
 ; TITLE OF INVENTION: O-Superfamily Conotoxin Peptides  
 ; FILE REFERENCE: 2314-227  
 ; CURRENT APPLICATION NUMBER: US/09/749,637A  
 ; CURRENT FILING DATE: 2000-12-28  
 ; PRIOR APPLICATION NUMBER: US 60/243,412  
 ; PRIOR FILING DATE: 2000-10-27  
 ; PRIOR APPLICATION NUMBER: US60/219,440  
 ; PRIOR FILING DATE: 2000-07-20  
 ; PRIOR APPLICATION NUMBER: US 60/214,263  
 ; PRIOR FILING DATE: 2000-06-26  
 ; PRIOR APPLICATION NUMBER: US 60/173,754

; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 409  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 92  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Conus omaria  
US-09-749-637A-92

Query Match 66.7%; Score 36; DB 10; Length 30;  
Best Local Similarity 75.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||| |||  
Db 5 CLDGGEIC 12

RESULT 8

US-09-749-637A-119  
; Sequence 119, Application US/09749637A  
; Patent No. US20020173449A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Utah Research Foundation  
; APPLICANT: Cognetix, Inc.  
; APPLICANT: Olivera, Baldomero M.  
; APPLICANT: Cartier, G. Edward  
; APPLICANT: Watkins, Maren  
; APPLICANT: Hillyard, David R.  
; APPLICANT: McIntosh, J. Michael  
; APPLICANT: Layer, Richard T.  
; APPLICANT: Jones, Robert M.  
; TITLE OF INVENTION: O-Superfamily Conotoxin Peptides  
; FILE REFERENCE: 2314-227  
; CURRENT APPLICATION NUMBER: US/09/749,637A  
; CURRENT FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/243,412  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US60/219,440  
; PRIOR FILING DATE: 2000-07-20  
; PRIOR APPLICATION NUMBER: US 60/214,263  
; PRIOR FILING DATE: 2000-06-26  
; PRIOR APPLICATION NUMBER: US 60/173,754  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 409  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 119  
; LENGTH: 77  
; TYPE: PRT  
; ORGANISM: Conus marmoreus  
US-09-749-637A-119

Query Match 66.7%; Score 36; DB 10; Length 77;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8

Db                    ||| | ||  
                      52 CLDGGEIC 59

RESULT 9

US-10-156-761-11025  
; Sequence 11025, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 11025  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-11025

Query Match                    66.7%; Score 36; DB 15; Length 219;  
Best Local Similarity       100.0%; Pred. No. 1.7e+02;  
Matches       5; Conservative       0; Mismatches       0; Indels       0; Gaps       0;

QY                    1 CLDWG 5  
                      |||||  
Db                    195 CLDWG 199

RESULT 10

US-09-749-637A-93  
; Sequence 93, Application US/09749637A  
; Patent No. US20020173449A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Utah Research Foundation  
; APPLICANT: Cognetix, Inc.  
; APPLICANT: Olivera, Baldomero M.  
; APPLICANT: Cartier, G. Edward  
; APPLICANT: Watkins, Maren  
; APPLICANT: Hillyard, David R.  
; APPLICANT: McIntosh, J. Michael  
; APPLICANT: Layer, Richard T.  
; APPLICANT: Jones, Robert M.  
; TITLE OF INVENTION: O-Superfamily Conotoxin Peptides  
; FILE REFERENCE: 2314-227  
; CURRENT APPLICATION NUMBER: US/09/749,637A



```

; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/243,412
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US60/219,440
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: US 60/214,263
; PRIOR FILING DATE: 2000-06-26
; PRIOR APPLICATION NUMBER: US 60/173,754
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 409
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 93
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Conus omaria
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(26)
; OTHER INFORMATION: Xaa at residue 6 may be Glu or gamma-carboxy-Glu; Xaa at
residue
; OTHER INFORMATION: 13 may be Pro or hydroxy-Pro; Xaa at residue 19 may be
Trp or bro
; OTHER INFORMATION: mo-Trp
US-09-749-637A-93

```

```

Query Match          64.8%; Score 35; DB 10; Length 26;
Best Local Similarity 75.0%; Pred. No. 38;
Matches      6; Conservative      0; Mismatches      2; Indels      0; Gaps      0;

```

```

Qy      1 CLDWGRIC 8
        ||| |||
Db      1 CLDGGXIC 8

```

# RESULT 11

```

US-09-749-637A-120
; Sequence 120, Application US/09749637A
; Patent No. US20020173449A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Watkins, Maren
; APPLICANT: Hillyard, David R.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Layer, Richard T.
; APPLICANT: Jones, Robert M.
; TITLE OF INVENTION: O-Superfamily Conotoxin Peptides
; FILE REFERENCE: 2314-227
; CURRENT APPLICATION NUMBER: US/09/749,637A
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/243,412
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US60/219,440
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: US 60/214,263

```

```

; PRIOR FILING DATE: 2000-06-26
; PRIOR APPLICATION NUMBER: US 60/173,754
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 409
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 120
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Conus marmoreus
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(26)
; OTHER INFORMATION: Xaa at residue 6 may be Glu or gamma-carboxy-Glu; Xaa at
residue
; OTHER INFORMATION: 13 may be Pro or hydroxy-Pro; Xaa at residue 19 may be
Trp or bro
; OTHER INFORMATION: mo-Trp
US-09-749-637A-120

```

```

Query Match          64.8%;  Score 35;  DB 10;  Length 26;
Best Local Similarity 75.0%;  Pred. No. 38;
Matches      6;  Conservative    0;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy      1 CLDWGRIC 8
        ||| |||
Db      1 CLDGGXIC 8

```

# RESULT 12

```

US-09-864-761-43385
; Sequence 43385, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43385
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003101.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EST_HUMAN HIT: BF035674.1, EVALUATE 5.00e-12
US-09-864-761-43385

```

```

Query Match          64.8%; Score 35; DB 9; Length 64;
Best Local Similarity 57.1%; Pred. No. 84;
Matches      4; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

```

```

Qy      1 CLDWGRI 7
        |: |||:
Db      54 CVQWGRV 60

```

# RESULT 13

```

US-09-925-299-996
; Sequence 996, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102

```

```

; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 996
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-996

```

```

Query Match          64.8%; Score 35; DB 9; Length 146;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches      5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLDWGR 6
        |||||
Db      28 CLDWNR 33

```

#### RESULT 14

```

US-09-925-299-996
; Sequence 996, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 996
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

```

; NAME/KEY: SITE  
; LOCATION: (14)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (16)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-996

Query Match 64.8%; Score 35; DB 11; Length 146;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
||| |  
Db 28 CLDWNR 33

RESULT 15  
US-09-738-626-5815  
; Sequence 5815, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5815  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5815

Query Match 64.8%; Score 35; DB 10; Length 381;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DWGRIC 8  
||| |  
Db 40 DWGSIC 45

Search completed: November 13, 2003, 09:58:28  
Job time : 16.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 8.33333 Seconds  
(without alignments)  
92.322 Million cell updates/sec

Title: US-09-228-866-8  
Perfect score: 54  
Sequence: 1 CLDWGRIC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	% Query Match	Length	DB	ID	Description
1	41	75.9	134	2	S28678	hypothetical prote	
2	41	75.9	247	1	WMVQ28	28K protein - pota	
3	41	75.9	247	2	S03546	hypothetical prote	
4	39	72.2	536	2	T24218	hypothetical prote	
5	39	72.2	690	2	T27357	hypothetical prote	
6	38	70.4	137	1	AYSP	acyl carrier prote	
7	38	70.4	339	1	TVRTM	protein kinase (EC	
8	38	70.4	418	2	A38197	protein kinase (EC	
9	37	68.5	64	2	C87540	hypothetical prote	

10	37	68.5	216	2	E69128	ribosomal protein
11	37	68.5	414	2	B96905	hypothetical prote
12	37	68.5	440	2	S14958	alpha-amylase (EC
13	36.5	67.6	418	2	B69360	asparaginase (asnA
14	36	66.7	304	2	G85068	N7-like protein [i
15	36	66.7	378	1	A40004	histidine decarbox
16	36	66.7	378	1	B40004	histidine decarbox
17	36	66.7	378	1	A25013	histidine decarbox
18	36	66.7	485	2	T06764	adenosylhomocystei
19	36	66.7	531	2	T04722	hypothetical prote
20	36	66.7	544	2	T13877	NADH2 dehydrogenas
21	36	66.7	576	2	T11046	NADH2 dehydrogenas
22	36	66.7	650	2	T36419	hypothetical prote
23	36	66.7	2180	2	T29764	hypothetical prote
24	35.5	65.7	206	2	AC1997	hypothetical prote
25	35	64.8	87	2	AH2124	transcription regu
26	35	64.8	89	2	S76057	hypothetical prote
27	35	64.8	174	2	S07146	gamma-s-crystallin
28	35	64.8	444	2	T26229	hypothetical prote
29	35	64.8	499	2	S28306	hypothetical prote
30	35	64.8	512	2	E69485	DNA-directed RNA p
31	35	64.8	523	2	S03572	DNA-directed RNA p
32	35	64.8	523	2	B84416	DNA-directed RNA p
33	35	64.8	579	2	D88551	protein T23G5.5 [i
34	35	64.8	615	2	T43330	catecholamine tran
35	35	64.8	642	2	C87410	iolC protein [impo
36	35	64.8	665	1	DIRTR1	protein-arginine d
37	35	64.8	673	1	DIMSR1	protein-arginine d
38	35	64.8	715	2	JC2222	major surface glyc
39	35	64.8	1076	2	JC2217	major surface glyc
40	35	64.8	1083	2	JC2300	cell surface glyco
41	35	64.8	1095	2	T41171	importin beta subu
42	35	64.8	1117	2	A75182	DNA-directed RNA p
43	35	64.8	1117	2	A71032	probable DNA-direc
44	35	64.8	1122	2	S25563	DNA-directed RNA p
45	35	64.8	1195	1	S26722	DNA-directed RNA p

#### ALIGNMENTS

RESULT 1  
 S28678  
 hypothetical protein 1 - phage SP01  
 C;Species: phage SP01  
 C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
 C;Accession: S28678  
 R;Scarlatto, V.; Sayre, M.H.  
 Gene 114, 115-119, 1992  
 A;Title: Sequence of the bacteriophage SP01 gene 30.  
 A;Reference number: S28678; MUID:92267370; PMID:1587473  
 A;Accession: S28678  
 A;Molecule type: DNA  
 A;Residues: 1-134 <SCA>  
 A;Cross-references: EMBL:M82842; NID:g216115; PIDN:AAA32596.1; PID:g216116  
 C;Genetics:  
 A;Start codon: GTG

Query Match 75.9%; Score 41; DB 2; Length 134;  
Best Local Similarity 71.4%; Pred. No. 6.1;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
|||:|  
Db 115 LDWGKVC 121

#### RESULT 2

WMVQ28

28K protein - potato leaf roll virus (strain 1)

C;Species: potato leaf roll virus

A;Note: host Solanum tuberosum (potato)

C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 21-Jul-2000

C;Accession: JA0117; S24590

R;Mayo, M.A.; Robinson, D.J.; Jolly, C.A.; Hyman, L.

J. Gen. Virol. 70, 1037-1051, 1989

A;Title: Nucleotide sequence of potato leafroll luteovirus RNA.

A;Reference number: JA0119; MUID:89279282; PMID:2732710

A;Accession: JA0117

A;Molecule type: genomic RNA

A;Residues: 1-247 <MAY>

A;Cross-references: EMBL:X14600; NID:g222293; PIDN:BAA00416.1; PID:g222296

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1989

C;Comment: The genome is a single-stranded, positive-sense RNA.

C;Superfamily: potato leaf roll virus 28K protein

Query Match 75.9%; Score 41; DB 1; Length 247;  
Best Local Similarity 62.5%; Pred. No. 10;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||:| :|  
Db 84 CLEWGLLC 91

#### RESULT 3

S03546

hypothetical protein 1 - potato leaf roll virus

C;Species: potato leaf roll virus

C;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 20-Sep-1999

C;Accession: S03546

R;van der Wilk, F.; Huisman, M.J.; Cornelissen, B.J.C.; Huttinga, H.; Goldbach, R.

FEBS Lett. 245, 51-56, 1989

A;Title: Nucleotide sequence and organization of potato leafroll virus genomic RNA.

A;Reference number: S03546; MUID:89171329; PMID:2466700

A;Accession: S03546

A;Molecule type: genomic RNA

A;Residues: 1-247 <VAN>

A;Cross-references: EMBL:Y07496; NID:g61198; PIDN:CAA68794.1; PID:g61199

C;Superfamily: potato leaf roll virus 28K protein



Query Match 75.9%; Score 41; DB 2; Length 247;  
Best Local Similarity 62.5%; Pred. No. 10;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||: || :|  
Db 84 CLEWGLLC 91

RESULT 4

T24218

hypothetical protein R13G10.2 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T24218

R;Gardner, A.

submitted to the EMBL Data Library, August 1994

A;Reference number: Z19857

A;Accession: T24218

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-536 <WIL>

A;Cross-references: EMBL:Z35602; PIDN:CAA84671.1; GSPDB:GN00021; CESP:R13G10.2

A;Experimental source: clone R13G10

C;Genetics:

A;Gene: CESP:R13G10.2

A;Map position: 3

A;Introns: 64/3; 194/1; 404/3

Query Match 72.2%; Score 39; DB 2; Length 536;  
Best Local Similarity 83.3%; Pred. No. 42;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
|: ||||  
Db 293 CIDWGR 298

RESULT 5

T27357

hypothetical protein Y70G10A.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T27357

R;Lloyd, C.

submitted to the EMBL Data Library, October 1998

A;Reference number: Z20354

A;Accession: T27357

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-690 <WIL>

A;Cross-references: EMBL:AL032660; PIDN:CAA21751.1; GSPDB:GN00021;

CESP:Y70G10A.3

A;Experimental source: clone Y70G10A

C;Genetics:

A;Gene: CESP:Y70G10A.3

A;Map position: 3

A;Introns: 61/3; 84/2; 185/1; 250/2; 326/3; 375/1; 398/3; 439/2; 490/3; 628/1; 655/1

Query Match 72.2%; Score 39; DB 2; Length 690;  
Best Local Similarity 62.5%; Pred. No. 52;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
||:| |  
Db 599 CLEWGESC 606

#### RESULT 6

AYSP

acyl carrier protein I precursor - spinach

C;Species: Spinacia oleracea (spinach)

C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 26-May-2000

C;Accession: A28052

R;Scherer, D.E.; Knauf, V.C.

Plant Mol. Biol. 9, 127-134, 1987

A;Title: Isolation of a cDNA clone for the acyl carrier protein-I of spinach.

A;Reference number: A28052

A;Accession: A28052

A;Molecule type: DNA

A;Residues: 1-137 <SCH>

C;Superfamily: acyl carrier protein; acyl carrier protein homology

C;Keywords: carrier protein; chloroplast; fatty acid biosynthesis;  
phosphopantetheine; phosphoprotein

F;1-56/Domain: transit peptide (chloroplast) #status predicted <TNP>

F;57-137/Product: acyl carrier protein #status predicted <MAT>

F;59-130/Domain: acyl carrier protein homology <ACP>

F;94/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.4%; Score 38; DB 1; Length 137;  
Best Local Similarity 83.3%; Pred. No. 20;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
|||||:  
Db 34 CLDWGK 39

#### RESULT 7

TVRTM

protein kinase (EC 2.7.1.37) MOS - rat

N;Alternate names: kinase-related transforming protein MOS; MOS proto-oncogene  
protein-serine/threonine kinase

C;Species: Rattus norvegicus (Norway rat)

C;Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999

C;Accession: A00648; I60596

R;Van der Hoorn, F.A.; Firzlaff, J.

Nucleic Acids Res. 12, 2147-2156, 1984

A;Title: Complete c-mos (rat) nucleotide sequence: presence of conserved domains  
in c-mos proteins.

A;Reference number: A00648; MUID:84144095; PMID:6322135

A;Accession: A00648

A;Molecule type: DNA

A;Residues: 1-339 <VAN>  
A;Note: the authors translated the codon TAC for residue 279 as His and GAG for 295 as Ala  
R;Leibovitch, S.A.; Lenormand, J.L.; Leibovitch, M.P.; Guiller, M.; Mallard, L.; Harel, J.  
Oncogene 5, 1149-1157, 1990  
A;Title: Rat myogenic c-mos cDNA: cloning sequence analysis and regulation during muscle development.  
A;Reference number: I60596; MUID:90363547; PMID:1697408  
A;Accession: I60596  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-46, 'V', 48-101, 'A', 103-294, 'A', 296-339 <RES>  
A;Cross-references: EMBL:X52952; NID:g55965; PIDN:CAA37128.1; PID:g55966  
C;Genetics:  
A;Gene: mos  
C;Superfamily: kinase-related transforming protein; protein kinase homology  
C;Keywords: ATP; phosphotransferase; proto-oncogene; serine/threonine-specific protein kinase  
F;59-338/Domain: protein kinase homology <KIN>  
F;67-75/Region: protein kinase ATP-binding motif  
F;88/Active site: Lys #status predicted

Query Match 70.4%; Score 38; DB 1; Length 339;  
Best Local Similarity 57.1%; Pred. No. 42;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
: |||::|  
Db 56 IDWGQVC 62

# RESULT 8

A38197  
protein kinase (EC 2.7.1.37) cdc2-like - human  
N;Alternate names: cholinesterase-related cell division control protein CHED  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 10-Sep-1997  
C;Accession: A38197  
R;Lapidot-Lifson, Y.; Patinkin, D.; Prody, C.A.; Ehrlich, G.; Seidman, S.; Ben-Aziz, R.; Benseler, F.; Eckstein, F.; Zakut, H.; Soreq, H.  
Proc. Natl. Acad. Sci. U.S.A. 89, 579-583, 1992  
A;Title: Cloning and antisense oligodeoxynucleotide inhibition of a human homolog of cdc2 required in hematopoiesis.  
A;Reference number: A38197; MUID:92115704; PMID:1731328  
A;Accession: A38197  
A;Molecule type: mRNA  
A;Residues: 1-418 <LAP>  
A;Note: sequence extracted from NCBI backbone (NCBIN:76015, NCBIP:76016)  
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology  
C;Keywords: ATP; cell cycle control; mitosis; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase  
F;89-353/Domain: protein kinase homology <KIN>  
F;97-105/Region: protein kinase ATP-binding motif  
F;101,257/Binding site: phosphate (Thr) (covalent) #status predicted  
F;102/Binding site: phosphate (Tyr) (covalent) #status predicted

F;120,223/Active site: Lys, Asp #status predicted

Query Match 70.4%; Score 38; DB 2; Length 418;  
Best Local Similarity 57.1%; Pred. No. 50;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LDWGRIC 8  
:||||:|  
Db 81 IDWGKLC 87

RESULT 9

C87540

hypothetical protein CC2348 [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: C87540

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: C87540

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-64 <STO>

A;Cross-references: GB:AE005673; NID:g13423875; PIDN:AAK24319.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2348

Query Match 68.5%; Score 37; DB 2; Length 64;  
Best Local Similarity 83.3%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
| ||||  
Db 36 CFDWGR 41

RESULT 10

E69128

ribosomal protein S5 - *Methanobacterium thermoautotrophicum* (strain Delta H)

N;Alternate names: eukaryotic ribosomal protein S2 homolog; prokaryotic ribosomal protein S5 homolog

C;Species: *Methanobacterium thermoautotrophicum*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 13-Aug-1999

C;Accession: E69128

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.;

Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.;  
Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta  
H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: E69128

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-216 <MTH>

A;Cross-references: GB:AE000796; GB:AE000666; NID:g2621057; PIDN:AAB84532.1;  
PID:g2621060

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH23

C;Superfamily: Escherichia coli ribosomal protein S5

Query Match 68.5%; Score 37; DB 2; Length 216;  
Best Local Similarity 62.5%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLDWGRIC 8  
| ||| :|  
Db 118 CGDWGCVC 125

#### RESULT 11

B96905

hypothetical protein CAC0042 [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C;Accession: B96905

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,  
R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hitti, J.; Wolf, Y.I.; Tatusov, R.L.;  
Sabathe, F.; Doucette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.;  
Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing  
Bacterium Clostridium acetobutylicum.

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: B96905

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-414 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK78029.1; PID:g15022864; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC0042

Query Match 68.5%; Score 37; DB 2; Length 414;  
Best Local Similarity 83.3%; Pred. No. 74;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLDWGR 6  
| |||| :  
Db 123 CLDWGQ 128

12	32	69.6	54	3	US-09-342-581-8	Sequence 8, Appli
13	32	69.6	92	2	US-07-728-215-41	Sequence 41, Appl
14	32	69.6	92	4	US-08-938-085A-41	Sequence 41, Appl
15	32	69.6	92	4	US-10-072-844-41	Sequence 41, Appl
16	32	69.6	131	1	US-08-757-541-10	Sequence 10, Appl
17	32	69.6	131	3	US-09-033-275-10	Sequence 10, Appl
18	32	69.6	131	3	US-09-342-581-10	Sequence 10, Appl
19	32	69.6	132	1	US-08-757-541-7	Sequence 7, Appli
20	32	69.6	132	1	US-08-757-541-11	Sequence 11, Appl
21	32	69.6	132	3	US-09-033-275-7	Sequence 7, Appli
22	32	69.6	132	3	US-09-033-275-11	Sequence 11, Appl
23	32	69.6	132	3	US-09-342-581-7	Sequence 7, Appli
24	32	69.6	132	3	US-09-342-581-11	Sequence 11, Appl
25	32	69.6	132	4	US-09-384-302A-10	Sequence 10, Appl
26	32	69.6	407	4	US-09-252-991A-29531	Sequence 29531, A
27	32	69.6	1037	4	US-09-340-620A-55	Sequence 55, Appl
28	32	69.6	1729	4	US-09-553-690-2	Sequence 2, Appli
29	31	67.4	8	4	US-09-187-859-2249	Sequence 2249, Ap
30	31	67.4	8	4	US-09-839-542B-2249	Sequence 2249, Ap
31	31	67.4	25	6	5447911-4	Patent No. 5447911
32	31	67.4	61	3	US-08-857-076-21	Sequence 21, Appl
33	31	67.4	92	4	US-08-938-085A-43	Sequence 43, Appl
34	31	67.4	92	4	US-10-072-844-43	Sequence 43, Appl
35	31	67.4	119	1	US-08-468-347-20	Sequence 20, Appl
36	31	67.4	119	1	US-08-226-264-24	Sequence 24, Appl
37	31	67.4	119	2	US-08-467-389-20	Sequence 20, Appl
38	31	67.4	119	2	US-08-779-379-20	Sequence 20, Appl
39	31	67.4	119	2	US-08-469-219-20	Sequence 20, Appl
40	31	67.4	119	3	US-09-228-152-19	Sequence 19, Appl
41	31	67.4	120	6	5447911-1	Patent No. 5447911
42	31	67.4	136	2	US-08-560-098A-59	Sequence 59, Appl
43	31	67.4	136	6	5189019-6	Patent No. 5189019
44	31	67.4	154	4	US-09-252-991A-21178	Sequence 21178, A
45	31	67.4	196	4	US-09-252-991A-21028	Sequence 21028, A

#### ALIGNMENTS

##### RESULT 1

US-08-526-710-9

; Sequence 9, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-9

```

```

Query Match          100.0%; Score 46; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CTRITESC 8
        |||||
Db      1 CTRITESC 8

```

# RESULT 2

```

US-08-862-855-9
; Sequence 9, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-9

```

```

Query Match          100.0%; Score 46; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CTRITESC 8
        |||||
Db      1 CTRITESC 8

```

# RESULT 3

US-09-226-985-9

```

; Sequence 9, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```



```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-9

```

```

Query Match          100.0%; Score 46; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

```

```

Qy      1 CTRITESC 8
        |||||
Db      1 CTRITESC 8

```

#### RESULT 4

```

US-09-227-906-9
; Sequence 9, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-9

```

```

Query Match          100.0%; Score 46; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      1 CTRITESC 8
        |||||
Db      1 CTRITESC 8

```

```

RESULT 5
US-09-860-793-5
; Sequence 5, Application US/09860793
; Patent No. 6559121
; GENERAL INFORMATION:
; APPLICANT: Pruett, John H
; APPLICANT: Temeyer, Kevin B
; APPLICANT: Kunz, Sidney E
; APPLICANT: Fisher, William F
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic
; TITLE OF INVENTION: Scabies
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.
; CURRENT APPLICATION NUMBER: US/09/860,793
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/366,603
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Psoroptes ovis
US-09-860-793-5

```

Query Match 71.7%; Score 33; DB 4; Length 90;  
Best Local Similarity 62.5%; Pred. No. 37;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||| | :|  
Db 70 CTRATRAC 77

#### RESULT 6

US-09-157-257-8

; Sequence 8, Application US/09157257  
; Patent No. 6375954  
; GENERAL INFORMATION:  
; APPLICANT: DUTTA, Sukanta K.  
; APPLICANT: BISWAS, Biswajit  
; APPLICANT: VEMULAPALLI, Ramesh  
; TITLE OF INVENTION: A SIZE-VARIABLE STRAIN-SPECIFIC PROTECTIVE ANTIGEN FOR  
; TITLE OF INVENTION: POTOMAC HORSE FEVER  
; FILE REFERENCE: 8172-9016  
; CURRENT APPLICATION NUMBER: US/09/157,257  
; CURRENT FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: 60/059,252  
; EARLIER FILING DATE: 1997-09-18  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Ehrlichia risticii  
US-09-157-257-8

Query Match 71.7%; Score 33; DB 4; Length 501;  
Best Local Similarity 75.0%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||| |||  
Db 480 CTRKKESC 487

#### RESULT 7

US-09-157-257-6

; Sequence 6, Application US/09157257  
; Patent No. 6375954  
; GENERAL INFORMATION:  
; APPLICANT: DUTTA, Sukanta K.  
; APPLICANT: BISWAS, Biswajit  
; APPLICANT: VEMULAPALLI, Ramesh  
; TITLE OF INVENTION: A SIZE-VARIABLE STRAIN-SPECIFIC PROTECTIVE ANTIGEN FOR  
; TITLE OF INVENTION: POTOMAC HORSE FEVER  
; FILE REFERENCE: 8172-9016  
; CURRENT APPLICATION NUMBER: US/09/157,257  
; CURRENT FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: 60/059,252  
; EARLIER FILING DATE: 1997-09-18

; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 539  
; TYPE: PRT  
; ORGANISM: Ehrlichia risticii  
US-09-157-257-6

Query Match 71.7%; Score 33; DB 4; Length 539;  
Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||| |||  
Db 518 CTRKKESC 525

RESULT 8

US-09-240-078-1  
; Sequence 1, Application US/09240078  
; Patent No. 6303749  
; GENERAL INFORMATION:  
; APPLICANT: Jarosinski, Mark A.  
; TITLE OF INVENTION: No. 6303749e1 Agouti and Agouti-Related Peptide Analogs  
; FILE REFERENCE: A-569  
; CURRENT APPLICATION NUMBER: US/09/240,078  
; CURRENT FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 48  
; TYPE: PRT  
; ORGANISM: Human  
US-09-240-078-1

Query Match 69.6%; Score 32; DB 4; Length 48;  
Best Local Similarity 62.5%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| |: |||  
Db 3 CVRLHESC 10

RESULT 9

US-09-031-902-2  
; Sequence 2, Application US/09031902  
; Patent No. 6228840  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Edward T.  
; APPLICANT: Quillan, J. Mark  
; APPLICANT: Sadee, Wolfgang  
; APPLICANT: Vlasov, Guennady  
; APPLICANT: Chang, J.K.  
; TITLE OF INVENTION: MELANOCORTIN RECEPTOR ANTAGONISTS AND  
; TITLE OF INVENTION: MODULATIONS OF MELANOCORTIN RECEPTOR ACTIVITY  
; NUMBER OF SEQUENCES: 12

```

;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Majestic, Parsons, Siebert & Hsue P.C.
;   STREET:    Four Embarcadero Center, Suite 1100
;   CITY:      San Francisco
;   STATE:     California
;   COUNTRY:   U.S.A.
;   ZIP:       94111-4106
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:    IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:    PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/09/031,902
;   FILING DATE:    27-FEB-1998
;   CLASSIFICATION:  514
;   ATTORNEY/AGENT INFORMATION:
;   NAME:          Siebert, J. Suzanne
;   REGISTRATION NUMBER:  28,758
;   REFERENCE/DOCKET NUMBER:  2500.095US0
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:     415-248-5500
;   TELEFAX:       415-362-5418
;   INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:       50 amino acids
;   TYPE:          amino acid
;   TOPOLOGY:     linear
;   MOLECULE TYPE:  protein
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:   NO
US-09-031-902-2

```

```

Query Match          69.6%;  Score 32;  DB 3;  Length 50;
Best Local Similarity 62.5%;  Pred. No. 32;
Matches      5;  Conservative    1;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy      1 CTRITESC 8
        | | : | | |
Db      5 CVRLHESC 12

```

# RESULT 10

US-08-757-541-8

```

; Sequence 8, Application US/08757541
; Patent No. 5766877
; GENERAL INFORMATION:
; APPLICANT:  Stark, Kevin Lee
; APPLICANT:  Luethy, Roland
; TITLE OF INVENTION:  NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES:  11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  AMGEN INC.
; STREET:    1840 DEHAVILLAND DRIVE
; CITY:      THOUSAND OAKS
; STATE:     CALIFORNIA
; COUNTRY:   USA

```

```

; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,541
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-402A
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-757-541-8

```

```

Query Match          69.6%; Score 32; DB 1; Length 54;
Best Local Similarity 62.5%; Pred. No. 34;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CTRITESC 8
        | | : |||
Db      9 CVRLHESC 16

```

# RESULT 11

```

US-09-033-275-8
; Sequence 8, Application US/09033275
; Patent No. 6060589
; GENERAL INFORMATION:
; APPLICANT: Stark, Kevin Lee
; APPLICANT: Luethy, Roland
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,275
; FILING DATE:
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-402A
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-033-275-8

```

```

Query Match          69.6%; Score 32; DB 3; Length 54;
Best Local Similarity 62.5%; Pred. No. 34;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CTRITESC 8
        | | : | | |
Db      9 CVRLHESC 16

```

RESULT 12

US-09-342-581-8

```

; Sequence 8, Application US/09342581
; Patent No. 6203995
; GENERAL INFORMATION:
; APPLICANT: Stark, Kevin Lee
; APPLICANT: Luethy, Roland
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/342,581
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/033,275
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-402A

```

; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-342-581-8

Query Match 69.6%; Score 32; DB 3; Length 54;  
Best Local Similarity 62.5%; Pred. No. 34;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | |  
Db 9 CVRLHESC 16

RESULT 13

US-07-728-215-41

; Sequence 41, Application US/07728215  
; Patent No. 5962643  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Dean  
; APPLICANT: Quaranta, Vito  
; APPLICANT: Pytela, Robert  
; TITLE OF INVENTION: A No. 5962643e1 Integrin Beta Subunit and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/728,215  
; FILING DATE: 19910711  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P31 8717  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 92 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear



; MOLECULE TYPE: protein  
US-07-728-215-41

Query Match 69.6%; Score 32; DB 2; Length 92;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| :|::|  
Db 55 CTLLTDTTC 62

RESULT 14

US-08-938-085A-41

; Sequence 41, Application US/08938085A

; Patent No. 6339148

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Dean

; APPLICANT: Quaranta, Vito

; APPLICANT: Pytela, Robert

; TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,085A

; FILING DATE: 26-SEP-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/728,215

; FILING DATE: 11-JUL-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Parent, Annette S.

; REGISTRATION NUMBER: 42,058

; REFERENCE/DOCKET NUMBER: 023070-080210US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 92 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-938-085A-41

Query Match 69.6%; Score 32; DB 4; Length 92;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| :|::|  
Db 55 CTTLTDTC 62

RESULT 15

US-10-072-844-41

; Sequence 41, Application US/10072844

; Patent No. 6576432

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Dean

; Quaranta, Vito

; Pytela, Robert

; TITLE OF INVENTION: A No. 6576432e1 Integrin Beta Subunit and Uses  
; Thereof

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/072,844

; FILING DATE: 06-Feb-2002

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,085A

; FILING DATE: 26-SEP-1997

; APPLICATION NUMBER: US 07/728,215

; FILING DATE: 11-JUL-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Parent, Annette S.

; REGISTRATION NUMBER: 42,058

; REFERENCE/DOCKET NUMBER: 023070-080210US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 92 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 41:

US-10-072-844-41

Query Match 69.6%; Score 32; DB 4; Length 92;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| :|::|  
Db 55 CTTLTDTC 62

Search completed: November 13, 2003, 09:54:59  
Job time : 10.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 26.9167 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-9  
Perfect score: 46  
Sequence: 1 CTRITESC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result	%		Query	Match	Length	DB	ID	Description
	No.	Score						
1	46	100.0	8	18	AAW13420		Brain homing pepti	
2	46	100.0	8	21	AAB07395		Brain homing pepti	
3	46	100.0	8	22	AAE11801		Phage peptide #9 t	
4	46	100.0	8	23	AAU10712		Brain homing pepti	
5	39	84.8	89	21	AAB40881		Human ORFX ORF645	
6	36	78.3	105	22	AAU41688		Propionibacterium	
7	35	76.1	53	24	ABB98725		Human PRiMA PRAD d	
8	35	76.1	53	24	ABB98733		Murine PRiMA PRAD	
9	35	76.1	78	24	ABB98728		Human PRiMA PRAD a	
10	35	76.1	78	24	ABB98736		Murine PRiMA PRAD	
11	35	76.1	88	24	ABB98731		Human PRiMA signal	
12	35	76.1	88	24	ABB98739		Murine PRiMA signa	
13	35	76.1	98	22	ABB63376		Drosophila melanog	
14	35	76.1	113	24	ABB98732		Human PRiMA signal	
15	35	76.1	113	24	ABB98740		Murine PRiMA signa	
16	35	76.1	118	24	ABB98730		Human PRiMA PRAD,	
17	35	76.1	118	24	ABB98738		Murine PRiMA PRAD,	
18	35	76.1	153	24	ABB98723		Human PRiMA. Homo	
19	35	76.1	153	24	ABB98724		Murine PRiMA. Mus	
20	35	76.1	3542	22	AAB62142		P. falciparum FCR3	
21	34	73.9	14	23	ABJ00587		B lymphocyte stimu	
22	34	73.9	14	23	ABG33448		B Lymphocyte Stimu	
23	34	73.9	72	21	AAG03340		Human secreted pro	
24	34	73.9	199	22	AAU51480		Propionibacterium	
25	33	71.7	78	22	AAG74005		Human colon cancer	
26	33	71.7	90	24	ABG72488		Modified Psoroptes	
27	33	71.7	278	23	ABB92220		Herbicidally activ	
28	33	71.7	297	21	AAG18109		Arabidopsis thalia	
29	33	71.7	299	21	AAG18108		Arabidopsis thalia	
30	33	71.7	306	21	AAG18107		Arabidopsis thalia	
31	33	71.7	453	21	AAB42692		Human ORFX ORF2456	
32	33	71.7	501	23	AAE23326		Ehrlichia risticii	
33	33	71.7	539	23	AAE23325		Ehrlichia risticii	
34	32	69.6	33	22	AAB75127		Human minimised ag	
35	32	69.6	33	23	AAU74943		Human minimised ag	
36	32	69.6	34	23	AAU74944		Human minimised ag	
37	32	69.6	34	23	AAU74945		Human minimised ag	
38	32	69.6	34	23	AAU74947		Human mini agouti	
39	32	69.6	46	20	AAAY49101		Human minimised ag	
40	32	69.6	46	20	AAAY49103		Mouse minimised ag	
41	32	69.6	46	22	AAB75126		Human minimised ag	
42	32	69.6	46	23	AAU74942		Human minimised ag	
43	32	69.6	48	21	AAB00081		Agouti related pep	
44	32	69.6	50	20	AAAY33951		Melanocortin-1 rec	
45	32	69.6	54	19	AAW26778		Human agouti-regul	

# ALIGNMENTS

RESULT 1

AAW13420

ID AAW13420 standard; Peptide; 8 AA.

XX

AC AAW13420;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;

KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in

PT vivo panning method, specifically to identify brain, kidney,

PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 15; Page 68; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
 CC peptide that was identified using a novel method for obtaining  
 CC molecules that home to a selected organ or tissue. This in vivo  
 CC panning method typically involves administering a phage display  
 CC library to a subject, and identifying expressed peptides which  
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
 CC vascular tissue or tumour tissue. The isolated peptides (see  
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
 CC labels to the selected organ/tissue (claimed) or to identify and/or  
 CC isolate target molecules (claimed). The peptides can be directly  
 CC identified in vivo, as compared to prior art in vitro screening  
 CC methods, which require further examination to see if they maintain  
 CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 46; DB 18; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8

Db                    |||||  
1 CTRITESC 8

RESULT 2

AAB07395

ID    AAB07395 standard; peptide; 8 AA.

XX

AC    AAB07395;

XX

DT    17-OCT-2000    (first entry)

XX

DE    Brain homing peptide # 9.

XX

KW    Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS    Mus sp.

XX

FH    Key                    Location/Qualifiers

FT    Disulfide-bond    1..8

FT                                /note= "Can optionally form a cyclic peptide"

XX

PN    US6068829-A.

XX

PD    30-MAY-2000.

XX

PF    23-JUN-1997;    97US-0862855.

XX

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Pasqualini R,   Ruoslahti E;

XX

DR    WPI; 2000-410850/35.

XX

PT    Identifying and recovering organ homing molecules or peptides by in  
PT    vivo panning comprises administering a library of diverse peptides  
PT    linked to a tag which facilitates recovery of these peptides -

XX

PS    Example 2; Column 17; 20pp; English.

XX

CC    The present sequence is a mouse brain homing peptide. This sequence was  
CC    identified by using in vivo panning to screen a library of potential  
CC    organ homing molecules. The present sequence can be used to direct a  
CC    moiety to a the brain tissue, by linking the moiety to the present  
CC    sequence. Examples of potential moieties are drugs, toxins or a  
CC    detectable label.

XX

SQ    Sequence    8 AA;

Query Match                    100.0%;    Score 46;    DB 21;    Length 8;

Best Local Similarity    100.0%;    Pred. No. 9.3e+05;

Matches    8;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy                    1 CTRITESC 8

Db

|||||||  
1 CTRITESC 8

RESULT 3

AAE11801

ID AAE11801 standard; peptide; 8 AA.

XX

AC AAE11801;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #9 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;  
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo  
PT panning that selectively home to a selected organ or tissue useful for  
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 46; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 CTRITESC 8  
|||  
Db 1 CTRITESC 8

RESULT 4

AAU10712

ID AAU10712 standard; peptide; 8 AA.

XX

AC AAU10712;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #9 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;

KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for

PT identifying molecules that home to a specific organ or tissue, e.g.

PT identifying a tumour homing molecule to identify the presence of cancer,

PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying

CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 46; DB 23; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|||  
Db 1 CTRITESC 8

RESULT 5

AAB40881

ID AAB40881 standard; Protein; 89 AA.

XX

AC AAB40881;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF645 polypeptide sequence SEQ ID NO:1290.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX

PA (CURA-) CURAGEN CORP.

XX  
 PI Shimkets RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC75090.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 1120; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 89 AA;

Query Match 84.8%; Score 39; DB 21; Length 89;  
 Best Local Similarity 75.0%; Pred. No. 15;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 |||: |||  
 Db 5 CTRVPESC 12

# RESULT 6

AAU41688

ID AAU41688 standard; Protein; 105 AA.

XX

AC AAU41688;

XX

DT 13-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #2584.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59515.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 2883; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 105 AA;

Query Match 78.3%; Score 36; DB 22; Length 105;  
 Best Local Similarity 75.0%; Pred. No. 58;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8

Db                    ||||| |  
66 CTRITRMC 73

RESULT 7

ABB98725

ID    ABB98725 standard; Protein; 53 AA.

XX

AC    ABB98725;

XX

DT    20-JAN-2003    (first entry)

XX

DE    Human PRiMA PRAD domain.

XX

KW    Human; PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;

KW    butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;

KW    cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;

KW    PRiMA-h.

XX

OS    Homo sapiens.

XX

PN    FR2822831-A1.

XX

PD    04-OCT-2002.

XX

PF    02-APR-2001; 2001FR-0004449.

XX

PR    02-APR-2001; 2001FR-0004449.

XX

PA    (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI    Krejci E,   Massoulie J,   Perrier A;

XX

DR    WPI; 2003-021468/02.

DR    N-PSDB; ABV74527.

XX

PT    New human protein that anchors cholinesterase to membranes and related

PT    nucleic acids, useful for treating disorders of cholinergic

PT    transmission, e.g. Alzheimer's disease -

XX

PS    Claim 3; Fig 3; 88pp; French.

XX

CC    The present invention relates to human and murine PRiMA (PRiMA-h or -s;

CC    see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially

CC    acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface

CC    membranes. Antibodies directed against PRiMA, and antisense

CC    oligonucleotides and mRNA directed against PRiMA coding sequence, are

CC    useful for treating diseases associated with reduction in levels of AChE,

CC    particularly disorders of cholinergic transmission either in central

CC    nervous system cells (particularly Alzheimer's diseases) or at the

CC    neuromuscular level (particularly myasthenia gravis). The present

CC    sequence is a PRiMA-h fragment.

XX

SQ    Sequence    53 AA;

Query Match                    76.1%;    Score 35;    DB 24;    Length 53;  
Best Local Similarity        50.0%;    Pred. No. 46;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|::|:|  
Db 6 CSKVTDSC 13

RESULT 8

ABB98733

ID ABB98733 standard; Protein; 53 AA.

XX

AC ABB98733;

XX

DT 20-JAN-2003 (first entry)

XX

DE Murine PRiMA PRAD domain.

XX

KW PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;  
KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;  
KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;  
KW murine; PRiMA-s.

XX

OS Mus musculus.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74535.

XX

PT New human protein that anchors cholinesterase to membranes and related  
PT nucleic acids, useful for treating disorders of cholinergic  
PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 3; Fig 11; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;  
CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially  
CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface  
CC membranes. Antibodies directed against PRiMA, and antisense  
CC oligonucleotides and mRNA directed against PRiMA coding sequence, are  
CC useful for treating diseases associated with reduction in levels of AChE,  
CC particularly disorders of cholinergic transmission either in central  
CC nervous system cells (particularly Alzheimer's diseases) or at the  
CC neuromuscular level (particularly myasthenia gravis). The present  
CC sequence is a PRiMA-s fragment.

XX

SQ Sequence 53 AA;

Query Match 76.1%; Score 35; DB 24; Length 53;  
Best Local Similarity 50.0%; Pred. No. 46;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|::|:|  
Db 6 CSKVTDSC 13

RESULT 9

ABB98728

ID ABB98728 standard; Protein; 78 AA.

XX

AC ABB98728;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human PRiMA PRAD and transmembrane domains.

XX

KW Human; PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;

KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;

KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;

KW PRiMA-h.

XX

OS Homo sapiens.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74530.

XX

PT New human protein that anchors cholinesterase to membranes and related

PT nucleic acids, useful for treating disorders of cholinergic

PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 3; Fig 6; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;

CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially

CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface

CC membranes. Antibodies directed against PRiMA, and antisense

CC oligonucleotides and mRNA directed against PRiMA coding sequence, are

CC useful for treating diseases associated with reduction in levels of AChE,

CC particularly disorders of cholinergic transmission either in central

CC nervous system cells (particularly Alzheimer's diseases) or at the

CC neuromuscular level (particularly myasthenia gravis). The present

CC sequence is a PRiMA-h fragment.

XX

SQ Sequence 78 AA;

Query Match 76.1%; Score 35; DB 24; Length 78;

Best Local Similarity 50.0%; Pred. No. 66;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8

|:::|:|

Db 6 CSKVTDSC 13

#### RESULT 10

ABB98736

ID ABB98736 standard; Protein; 78 AA.

XX

AC ABB98736;

XX

DT 20-JAN-2003 (first entry)

XX

DE Murine PRiMA PRAD and transmembrane domains.

XX

KW PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;

KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;

KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;

KW murine; PRiMA-s.

XX

OS Mus musculus.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74538.

XX

PT New human protein that anchors cholinesterase to membranes and related

PT nucleic acids, useful for treating disorders of cholinergic

PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 6; Fig 14; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;

CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially

CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface

CC membranes. Antibodies directed against PRiMA, and antisense

CC oligonucleotides and mRNA directed against PRiMA coding sequence, are

CC useful for treating diseases associated with reduction in levels of AChE,



CC particularly disorders of cholinergic transmission either in central  
CC nervous system cells (particularly Alzheimer's diseases) or at the  
CC neuromuscular level (particularly myasthenia gravis). The present  
CC sequence is a PRiMA-s fragment.

XX

SQ Sequence 78 AA;

Query Match 76.1%; Score 35; DB 24; Length 78;  
Best Local Similarity 50.0%; Pred. No. 66;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|::|:|  
Db 6 CSKVTDSC 13

RESULT 11

ABB98731

ID ABB98731 standard; Protein; 88 AA.

XX

AC ABB98731;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human PRiMA signal peptide and PRAD domain.

XX

KW Human; PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;  
KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;  
KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;  
KW PRiMA-h.

XX

OS Homo sapiens.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74533.

XX

PT New human protein that anchors cholinesterase to membranes and related  
PT nucleic acids, useful for treating disorders of cholinergic  
PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 3; Fig 9; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;  
CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially  
CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface

CC membranes. Antibodies directed against PRiMA, and antisense  
CC oligonucleotides and mRNA directed against PRiMA coding sequence, are  
CC useful for treating diseases associated with reduction in levels of AChE,  
CC particularly disorders of cholinergic transmission either in central  
CC nervous system cells (particularly Alzheimer's diseases) or at the  
CC neuromuscular level (particularly myasthenia gravis). The present  
CC sequence is a PRiMA-h fragment.

XX

SQ Sequence 88 AA;

Query Match 76.1%; Score 35; DB 24; Length 88;  
Best Local Similarity 50.0%; Pred. No. 74;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|:::|:|  
Db 41 CSKVTDSC 48

#### RESULT 12

ABB98739

ID ABB98739 standard; Protein; 88 AA.

XX

AC ABB98739;

XX

DT 20-JAN-2003 (first entry)

XX

DE Murine PRiMA signal peptide and PRAD domain.

XX

KW PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;  
KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;  
KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;  
KW murine; PRiMA-s.

XX

OS Mus musculus.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74541.

XX

PT New human protein that anchors cholinesterase to membranes and related  
PT nucleic acids, useful for treating disorders of cholinergic  
PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 3; Fig 17; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;  
CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially  
CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface  
CC membranes. Antibodies directed against PRiMA, and antisense  
CC oligonucleotides and mRNA directed against PRiMA coding sequence, are  
CC useful for treating diseases associated with reduction in levels of AChE,  
CC particularly disorders of cholinergic transmission either in central  
CC nervous system cells (particularly Alzheimer's diseases) or at the  
CC neuromuscular level (particularly myasthenia gravis). The present  
CC sequence is a PRiMA-s fragment.

XX

SQ Sequence 88 AA;

Query Match 76.1%; Score 35; DB 24; Length 88;  
Best Local Similarity 50.0%; Pred. No. 74;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|::|:|  
Db 41 CSKVTDESC 48

#### RESULT 13

ABB63376

ID ABB63376 standard; Protein; 98 AA.

XX

AC ABB63376;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 16920.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL07479.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX

PS Disclosure; SEQ ID NO 16920; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 98 AA;

Query Match 76.1%; Score 35; DB 22; Length 98;  
 Best Local Similarity 62.5%; Pred. No. 82;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 |||: |:|  
 Db 25 CTRLRENC 32

#### RESULT 14

ABB98732

ID ABB98732 standard; Protein; 113 AA.

XX

AC ABB98732;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human PRiMA signal peptide, PRAD & transmembrane domain.

XX

KW Human; PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;  
 KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;  
 KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;  
 KW PRiMA-h.

XX

OS Homo sapiens.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.

DR N-PSDB; ABV74534.

XX

PT New human protein that anchors cholinesterase to membranes and related  
PT nucleic acids, useful for treating disorders of cholinergic  
PT transmission, e.g. Alzheimer's disease -

XX

PS Claim 3; Fig 10; 88pp; French.

XX

CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;  
CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially  
CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface  
CC membranes. Antibodies directed against PRiMA, and antisense  
CC oligonucleotides and mRNA directed against PRiMA coding sequence, are  
CC useful for treating diseases associated with reduction in levels of AChE,  
CC particularly disorders of cholinergic transmission either in central  
CC nervous system cells (particularly Alzheimer's diseases) or at the  
CC neuromuscular level (particularly myasthenia gravis). The present  
CC sequence is a PRiMA-h fragment.

XX

SQ Sequence 113 AA;

Query Match 76.1%; Score 35; DB 24; Length 113;  
Best Local Similarity 50.0%; Pred. No. 93;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|:::|:|  
Db 41 CSKVTDS 48

#### RESULT 15

ABB98740

ID ABB98740 standard; Protein; 113 AA.

XX

AC ABB98740;

XX

DT 20-JAN-2003 (first entry)

XX

DE Murine PRiMA signal peptide, PRAD & transmembrane domain.

XX

KW PRiMA; cholinesterase; acetylcholinesterase; myasthenia gravis;  
KW butyrylcholinesterase; AChE; BChE; cell-surface membrane; nootropic;  
KW cholinergic transmission; Alzheimer's disease; anchor; neuroprotective;  
KW murine; PRiMA-s.

XX

OS Mus musculus.

XX

PN FR2822831-A1.

XX

PD 04-OCT-2002.

XX

PF 02-APR-2001; 2001FR-0004449.

XX

PR 02-APR-2001; 2001FR-0004449.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Krejci E, Massoulie J, Perrier A;

XX

DR WPI; 2003-021468/02.  
 DR N-PSDB; ABV74542.  
 XX  
 PT New human protein that anchors cholinesterase to membranes and related  
 PT nucleic acids, useful for treating disorders of cholinergic  
 PT transmission, e.g. Alzheimer's disease -  
 XX  
 PS Claim 3; Fig 18; 88pp; French.  
 XX  
 CC The present invention relates to human and murine PRiMA (PRiMA-h or -s;  
 CC see ABB98723 and ABB98724). PRiMA anchors cholinesterases, especially  
 CC acetyl- or butyryl-cholinesterases (AChE or BChE), to cell-surface  
 CC membranes. Antibodies directed against PRiMA, and antisense  
 CC oligonucleotides and mRNA directed against PRiMA coding sequence, are  
 CC useful for treating diseases associated with reduction in levels of AChE,  
 CC particularly disorders of cholinergic transmission either in central  
 CC nervous system cells (particularly Alzheimer's diseases) or at the  
 CC neuromuscular level (particularly myasthenia gravis). The present  
 CC sequence is a PRiMA-s fragment.  
 XX  
 SQ Sequence 113 AA;

Query Match 76.1%; Score 35; DB 24; Length 113;  
 Best Local Similarity 50.0%; Pred. No. 93;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 |:::||  
 Db 41 CSKVTDSC 48

Search completed: November 13, 2003, 09:45:28  
 Job time : 27.9167 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 16.5833 Seconds  
 (without alignments)  
 88.069 Million cell updates/sec

Title: US-09-228-866-9  
 Perfect score: 46  
 Sequence: 1 CTRITESC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	34	73.9	14	11	US-09-932-613-45	Sequence 45, Appl
2	34	73.9	14	12	US-09-932-322-45	Sequence 45, Appl
3	34	73.9	586	15	US-10-156-761-12890	Sequence 12890, A
4	33	71.7	78	15	US-10-106-698-4779	Sequence 4779, Ap
5	33	71.7	90	10	US-09-860-793-5	Sequence 5, Appli
6	33	71.7	804	15	US-10-156-761-7708	Sequence 7708, Ap
7	32	69.6	54	9	US-09-754-862-8	Sequence 8, Appli
8	32	69.6	54	15	US-10-256-590-8	Sequence 8, Appli
9	32	69.6	92	14	US-10-072-841-41	Sequence 41, Appl
10	32	69.6	131	9	US-09-754-862-10	Sequence 10, Appl
11	32	69.6	131	15	US-10-256-590-10	Sequence 10, Appl
12	32	69.6	132	9	US-09-754-862-7	Sequence 7, Appli
13	32	69.6	132	9	US-09-754-862-11	Sequence 11, Appl
14	32	69.6	132	15	US-10-256-590-7	Sequence 7, Appli
15	32	69.6	132	15	US-10-256-590-11	Sequence 11, Appl
16	32	69.6	324	12	US-10-017-161-54	Sequence 54, Appl
17	32	69.6	343	12	US-10-017-161-804	Sequence 804, App
18	32	69.6	1037	9	US-09-728-721-55	Sequence 55, Appl
19	32	69.6	1037	15	US-10-295-981-55	Sequence 55, Appl
20	32	69.6	1729	12	US-09-840-743-2	Sequence 2, Appli
21	31	67.4	8	15	US-10-006-869-2249	Sequence 2249, Ap
22	31	67.4	29	9	US-09-904-380-28	Sequence 28, Appl
23	31	67.4	57	9	US-09-864-761-40354	Sequence 40354, A

24	31	67.4	61	9	US-09-205-658-21	Sequence 21, Appl
25	31	67.4	61	9	US-09-844-353A-21	Sequence 21, Appl
26	31	67.4	61	12	US-09-963-693-21	Sequence 21, Appl
27	31	67.4	65	9	US-09-864-761-39420	Sequence 39420, A
28	31	67.4	132	10	US-09-731-872-297	Sequence 297, App
29	31	67.4	132	12	US-09-876-997-297	Sequence 297, App
30	31	67.4	437	14	US-10-042-417-54	Sequence 54, Appl
31	31	67.4	769	14	US-10-072-841-31	Sequence 31, Appl
32	31	67.4	788	14	US-10-072-841-27	Sequence 27, Appl
33	31	67.4	788	15	US-10-171-311-101	Sequence 101, App
34	31	67.4	2139	9	US-09-727-384-6	Sequence 6, Appli
35	31	67.4	2139	15	US-10-023-219-4	Sequence 4, Appli
36	30	65.2	52	12	US-10-080-254-64	Sequence 64, Appl
37	30	65.2	86	15	US-10-106-698-6220	Sequence 6220, Ap
38	30	65.2	93	15	US-10-128-714-3382	Sequence 3382, Ap
39	30	65.2	155	9	US-09-925-301-1337	Sequence 1337, Ap
40	30	65.2	170	10	US-09-738-626-6681	Sequence 6681, Ap
41	30	65.2	293	12	US-10-017-161-818	Sequence 818, App
42	30	65.2	349	10	US-09-976-736-8	Sequence 8, Appli
43	30	65.2	349	11	US-09-972-473-17	Sequence 17, Appl
44	30	65.2	350	10	US-09-909-320-236	Sequence 236, App
45	30	65.2	350	10	US-09-909-088B-236	Sequence 236, App

#### ALIGNMENTS

##### RESULT 1

US-09-932-613-45

```
; Sequence 45, Application US/09932613
; Publication No. US20030091565A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; APPLICANT: Beltzer, James P.
; APPLICANT: Potter, M. Daniel
; APPLICANT: Fleming, Tony J.
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: BINDING POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: Dyx-025.1 PCT; DYX-025.1 US
; CURRENT APPLICATION NUMBER: US/09/932,613
; CURRENT FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 458
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: BLYS binding polypeptide
US-09-932-613-45
```

```
Query Match          73.9%;  Score 34;  DB 11;  Length 14;
Best Local Similarity 62.5%;  Pred. No. 8.3;
Matches      5;  Conservative      2;  Mismatches      1;  Indels      0;  Gaps      0;
```

```
Qy      1 CTRITESC 8
        | |:|:|
```



Db

4 CDRLTKSC 11

RESULT 2

US-09-932-322-45

; Sequence 45, Application US/09932322  
; Publication No. US20030194743A1  
; GENERAL INFORMATION:  
; APPLICANT: Dyax Corp.  
; APPLICANT: Beltzer, James P.  
; APPLICANT: Potter, M. Daniel  
; APPLICANT: Fleming, Tony J.  
; APPLICANT: Ladner, Robert Charles  
; TITLE OF INVENTION: BINDING POLYPEPTIDES FOR B LYMPHOCYTE STIMULATOR PROTEIN (BLyS)  
; FILE REFERENCE: Dyx-018.1 PCT; DYX-018.1 US  
; CURRENT APPLICATION NUMBER: US/09/932,322  
; CURRENT FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 458  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 45  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: BLyS binding polypeptide  
US-09-932-322-45

Query Match 73.9%; Score 34; DB 12; Length 14;  
Best Local Similarity 62.5%; Pred. No. 8.3;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : ||  
Db 4 CDRLTKSC 11

RESULT 3

US-10-156-761-12890

; Sequence 12890, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 12890  
; LENGTH: 586  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-12890

Query Match 73.9%; Score 34; DB 15; Length 586;  
Best Local Similarity 62.5%; Pred. No. 3.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|||:| :|  
Db 164 CTRLTGTC 171

RESULT 4

US-10-106-698-4779  
; Sequence 4779, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and  
Polypeptides  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 4779  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-106-698-4779

Query Match 71.7%; Score 33; DB 15; Length 78;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| :| |||  
Db 2 CNKIVESC 9

RESULT 5

US-09-860-793-5  
; Sequence 5, Application US/09860793  
; Patent No. US20020136734A1  
; GENERAL INFORMATION:  
; APPLICANT: Pruett, John H  
; APPLICANT: Temeyer, Kevin B

; APPLICANT: Kunz, Sidney E  
; APPLICANT: Fisher, William F  
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic  
; TITLE OF INVENTION: Scabies  
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.  
; CURRENT APPLICATION NUMBER: US/09/860,793  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/366,603  
; PRIOR FILING DATE: 1999-08-03  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 90  
; TYPE: PRT  
; ORGANISM: Psoroptes ovis  
US-09-860-793-5

Query Match 71.7%; Score 33; DB 10; Length 90;  
Best Local Similarity 62.5%; Pred. No. 77;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||| | :|  
Db 70 CTRATRAC 77

#### RESULT 6

US-10-156-761-7708  
; Sequence 7708, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7708  
; LENGTH: 804  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7708

Query Match 71.7%; Score 33; DB 15; Length 804;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | | : | |  
Db 353 CTRLREPC 360

RESULT 7

US-09-754-862-8

; Sequence 8, Application US/09754862  
; Patent No. US20010007752A1  
; GENERAL INFORMATION:  
; APPLICANT: Stark, Kevin Lee  
; APPLICANT: Luethy, Roland  
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/754,862  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/342,581  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OLESKI, NANCY A  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-402A  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-754-862-8

Query Match 69.6%; Score 32; DB 9; Length 54;  
Best Local Similarity 62.5%; Pred. No. 71;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | |  
Db 9 CVRLHESC 16

RESULT 8

US-10-256-590-8

```

; Sequence 8, Application US/10256590
; Publication No. US20030082737A1
; GENERAL INFORMATION:
;   APPLICANT: Stark, Kevin Lee
;             Luethy, Roland
;   TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: AMGEN INC.
;     STREET: 1840 DEHAVILLAND DRIVE
;     CITY: THOUSAND OAKS
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 91320-1789
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/256,590
;     FILING DATE: 27-Sep-2002
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/09/754,862
;     FILING DATE: <Unknown>
;     APPLICATION NUMBER: 09/342,581
;     FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;     NAME: OLESKI, NANCY A
;     REGISTRATION NUMBER: 34,688
;     REFERENCE/DOCKET NUMBER: A-402A
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 54 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-256-590-8

```

```

Query Match          69.6%;  Score 32;  DB 15;  Length 54;
Best Local Similarity 62.5%;  Pred. No. 71;
Matches      5;  Conservative    1;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy      1 CTRITESC 8
        | | : |||
Db      9 CVRLHESC 16

```

```

RESULT 9
US-10-072-841-41
; Sequence 41, Application US/10072841
; Publication No. US20020164708A1
; GENERAL INFORMATION:
;   APPLICANT: Sheppard, Dean

```

```

;           Quaranta, Vito
;           Pytela, Robert
;   TITLE OF INVENTION: A No. US20020164708A1e1 Integrin Beta Subunit and
Uses
;
;           Thereof
;   NUMBER OF SEQUENCES: 43
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
;       STREET: 4370 La Jolla Village Drive, Suite 700
;       CITY: San Diego
;       STATE: California
;       COUNTRY: United States of America
;       ZIP: 92122
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/072,841
;       FILING DATE: 06-Feb-2002
;       CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/728,215
;       FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Campbell, Cathryn A.
;       REGISTRATION NUMBER: 31,815
;       REFERENCE/DOCKET NUMBER: P31 8717
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (619) 535-9001
;       TELEFAX: (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 41:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 92 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-10-072-841-41

```

```

Query Match          69.6%;  Score 32;  DB 14;  Length 92;
Best Local Similarity 50.0%;  Pred. No. 1.2e+02;
Matches      4;  Conservative    3;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy      1 CTRITESC 8
        || :|::|
Db      55 CTTLTDTC 62

```

```

RESULT 10
US-09-754-862-10
; Sequence 10, Application US/09754862
; Patent No. US20010007752A1
;   GENERAL INFORMATION:
;       APPLICANT: Stark, Kevin Lee
;       APPLICANT: Luethy, Roland

```

```

; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/754,862
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/342,581
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-402A
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-754-862-10

```

```

Query Match          69.6%; Score 32; DB 9; Length 131;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CTRITESC 8
        | | : |||
Db      86 CVRLHESC 93

```

```

RESULT 11
US-10-256-590-10
; Sequence 10, Application US/10256590
; Publication No. US20030082737A1
; GENERAL INFORMATION:
; APPLICANT: Stark, Kevin Lee
;           Luethy, Roland
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CALIFORNIA

```

```

;          COUNTRY: USA
;          ZIP: 91320-1789
;    COMPUTER READABLE FORM:
;          MEDIUM TYPE: Floppy disk
;          COMPUTER: IBM PC compatible
;          OPERATING SYSTEM: PC-DOS/MS-DOS
;          SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/10/256,590
;          FILING DATE: 27-Sep-2002
;          CLASSIFICATION: <Unknown>
;    PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: US/09/754,862
;          FILING DATE: <Unknown>
;          APPLICATION NUMBER: 09/342,581
;          FILING DATE: <Unknown>
;    ATTORNEY/AGENT INFORMATION:
;          NAME: OLESKI, NANCY A
;          REGISTRATION NUMBER: 34,688
;          REFERENCE/DOCKET NUMBER: A-402A
;    INFORMATION FOR SEQ ID NO: 10:
;          SEQUENCE CHARACTERISTICS:
;            LENGTH: 131 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-256-590-10

```

```

Query Match          69.6%;  Score 32;  DB 15;  Length 131;
Best Local Similarity 62.5%;  Pred. No. 1.7e+02;
Matches      5;  Conservative    1;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy          1 CTRITESC 8
             | | : | | |
Db          86 CVRLHESC 93

```

# RESULT 12

US-09-754-862-7

```

; Sequence 7, Application US/09754862
; Patent No. US20010007752A1
;  GENERAL INFORMATION:
;    APPLICANT: Stark, Kevin Lee
;    APPLICANT: Luethy, Roland
;    TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
;    NUMBER OF SEQUENCES: 11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: AMGEN INC.
;      STREET: 1840 DEHAVILLAND DRIVE
;      CITY: THOUSAND OAKS
;      STATE: CALIFORNIA
;      COUNTRY: USA
;      ZIP: 91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk

```



```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/754,862
;      FILING DATE:
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/342,581
;      FILING DATE:
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  OLESKI, NANCY A
;      REGISTRATION NUMBER:  34,688
;      REFERENCE/DOCKET NUMBER:  A-402A
;      INFORMATION FOR SEQ ID NO:  7:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-754-862-7

```

```

Query Match          69.6%;  Score 32;  DB 9;  Length 132;
Best Local Similarity 62.5%;  Pred. No. 1.7e+02;
Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

```

```

Qy      1 CTRITESC 8
        | | : | | |
Db      87 CVRLHESC 94

```

# RESULT 13

```

US-09-754-862-11
; Sequence 11, Application US/09754862
; Patent No. US20010007752A1
; GENERAL INFORMATION:
; APPLICANT:  Stark, Kevin Lee
; APPLICANT:  Luethy, Roland
; TITLE OF INVENTION:  NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES:  11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  AMGEN INC.
; STREET:  1840 DEHAVILLAND DRIVE
; CITY:  THOUSAND OAKS
; STATE:  CALIFORNIA
; COUNTRY:  USA
; ZIP:  91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/754,862
; FILING DATE:
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/342,581
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-402A
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-754-862-11

```

```

Query Match          69.6%; Score 32; DB 9; Length 132;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CTRITESC 8
        | |: |||
Db      87 CVRLHESC 94

```

#### RESULT 14

US-10-256-590-7

```

; Sequence 7, Application US/10256590
; Publication No. US20030082737A1
; GENERAL INFORMATION:
; APPLICANT: Stark, Kevin Lee
;           Luethy, Roland
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/256,590
; FILING DATE: 27-Sep-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/754,862
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 09/342,581
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OLESKI, NANCY A

```

;           REGISTRATION NUMBER: 34,688  
;           REFERENCE/DOCKET NUMBER: A-402A  
;   INFORMATION FOR SEQ ID NO: 7:  
;       SEQUENCE CHARACTERISTICS:  
;           LENGTH: 132 amino acids  
;           TYPE: amino acid  
;           STRANDEDNESS: single  
;           TOPOLOGY: linear  
;       MOLECULE TYPE: protein  
;       SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-256-590-7

Query Match           69.6%;   Score 32;   DB 15;   Length 132;  
Best Local Similarity   62.5%;   Pred. No. 1.7e+02;  
Matches    5;   Conservative    1;   Mismatches    2;   Indels    0;   Gaps    0;

Qy           1 CTRITESC 8  
             | | : | | |  
Db           87 CVRLHESC 94

RESULT 15

US-10-256-590-11

; Sequence 11, Application US/10256590  
; Publication No. US20030082737A1  
;   GENERAL INFORMATION:  
;       APPLICANT: Stark, Kevin Lee  
;               Luethy, Roland  
;       TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE  
;       NUMBER OF SEQUENCES: 11  
;       CORRESPONDENCE ADDRESS:  
;           ADDRESSEE: AMGEN INC.  
;           STREET: 1840 DEHAVILLAND DRIVE  
;           CITY: THOUSAND OAKS  
;           STATE: CALIFORNIA  
;           COUNTRY: USA  
;           ZIP: 91320-1789  
;       COMPUTER READABLE FORM:  
;           MEDIUM TYPE: Floppy disk  
;           COMPUTER: IBM PC compatible  
;           OPERATING SYSTEM: PC-DOS/MS-DOS  
;           SOFTWARE: PatentIn Release #1.0, Version #1.30  
;       CURRENT APPLICATION DATA:  
;           APPLICATION NUMBER: US/10/256,590  
;           FILING DATE: 27-Sep-2002  
;           CLASSIFICATION: <Unknown>  
;       PRIOR APPLICATION DATA:  
;           APPLICATION NUMBER: US/09/754,862  
;           FILING DATE: <Unknown>  
;           APPLICATION NUMBER: 09/342,581  
;           FILING DATE: <Unknown>  
;       ATTORNEY/AGENT INFORMATION:  
;           NAME: OLESKI, NANCY A  
;           REGISTRATION NUMBER: 34,688  
;           REFERENCE/DOCKET NUMBER: A-402A  
;   INFORMATION FOR SEQ ID NO: 11:  
;       SEQUENCE CHARACTERISTICS:



## SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	35	76.1	98	2	S62338	L71-10 protein - f
2	35	76.1	186	2	E82625	outer membrane pro
3	34	73.9	237	2	F95411	hypothetical prote
4	34	73.9	317	2	B86941	hypothetical prote
5	34	73.9	319	2	E70623	hypothetical prote
6	33	71.7	5825	2	T12117	polyprotein - fava
7	32	69.6	92	2	D37057	epithelial cell gl
8	32	69.6	289	2	S77232	hypothetical prote
9	32	69.6	306	2	T39702	probable peroxisom
10	32	69.6	366	2	C88940	protein C05E4.12 [
11	32	69.6	443	2	A70438	flagellar export p
12	32	69.6	741	2	T00206	epidermis-specific
13	31	67.4	32	2	B44007	aptotoxin VII - tr
14	31	67.4	101	2	JQ2199	UL50h protein - Ma
15	31	67.4	119	2	A34398	antistasin - Mexic
16	31	67.4	119	2	A34816	anticoagulant prot
17	31	67.4	119	2	S13904	antistasin - Mexic
18	31	67.4	120	2	B28806	antistasin B - Mex
19	31	67.4	120	2	A28806	antistasin A - Mex
20	31	67.4	136	2	JS0209	antistasin precurs
21	31	67.4	178	2	A23219	high-cysteine chor
22	31	67.4	245	2	S48363	hypothetical prote
23	31	67.4	312	2	S51085	CdK-activating kin
24	31	67.4	545	2	T14632	hypothetical prote
25	31	67.4	744	2	S38279	cell surface recep
26	31	67.4	744	2	D88462	protein daf-4 [imp
27	31	67.4	747	2	F69332	heterodisulfide re
28	31	67.4	760	2	S19374	probable membrane
29	31	67.4	769	1	IJHULM	leukocyte adhesion
30	31	67.4	770	2	I40027	nitrite reductase
31	31	67.4	788	2	A37057	integrin beta-6 ch
32	31	67.4	984	2	T50309	hypothetical WD-re
33	31	67.4	1360	2	F96596	hypothetical prote
34	31	67.4	1652	2	T16799	hypothetical prote
35	31	67.4	1817	2	T34249	hypothetical prote
36	31	67.4	2168	2	T30171	ninein - mouse
37	31	67.4	5376	2	T42215	zonadhesin - mouse
38	30	65.2	108	2	AG2472	hypothetical prote
39	30	65.2	120	2	A70382	hypothetical prote
40	30	65.2	144	2	E88485	protein F23F12.3 [
41	30	65.2	203	2	C64892	probable CDP-alcoh
42	30	65.2	203	2	G85738	probable enzyme Z2
43	30	65.2	203	2	B90880	probable enzyme [i
44	30	65.2	210	2	S53300	self-incompatibili
45	30	65.2	222	2	JQ1074	S1 protein - garde

## ALIGNMENTS

RESULT 1

S62338

L71-10 protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 24-Sep-1999

C;Accession: S62338; S62348

R;Wright, L.G.; Chen, T.; Thummel, C.S.; Guild, G.M.

J. Mol. Biol. 255, 387-400, 1996

A;Title: Molecular characterization of the 71E late puff in *Drosophila melanogaster* reveals a family of novel genes.

A;Reference number: S62333; MUID:96152797; PMID:8568884

A;Accession: S62338

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-98 <WRI>

A;Cross-references: EMBL:U23836

A;Accession: S62348

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-98 <WRW>

A;Cross-references: EMBL:U24574; NID:g775244; PIDN:AAA65118.1; PID:g775245

C;Genetics:

A;Gene: L71-10

A;Cross-references: FlyBase:FBgn0014850

A;Introns: 78/1

C;Superfamily: L71-10 protein

Query Match	76.1%;	Score 35;	DB 2;	Length 98;
Best Local Similarity	62.5%;	Pred. No. 14;		
Matches	5;	Conservative	2;	Mismatches 1; Indels 0; Gaps 0;

Qy	1	CTRITESC	8
		:  :	
Db	25	CTRLRENC	32

## RESULT 2

E82625

outer membrane protein P6 precursor XF1896 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C;Accession: E82625

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82625

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-186 <SIM>

A;Cross-references: GB:AE004009; GB:AE003849; NID:g9106980; PIDN:AAF84702.1;

GSPDB:GN00128; XFSC:XF1896

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.;

Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros,

M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF1896

Query Match 76.1%; Score 35; DB 2; Length 186;  
Best Local Similarity 75.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| |||  
Db 165 CTESTESC 172

### RESULT 3

F95411

hypothetical protein Sma2221 [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C;Accession: F95411

R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowser, L.; Capela, D.; Galibert, F.; Gouzy, J.; Gurjal, M.; Hong, A.; Huizar, L.; Hyman, R.W.; Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.; Davis, R.W.; Federspiel, N.A.; Long, S.R.

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid.

A;Reference number: A95262; MUID:21396509; PMID:11481432

A;Accession: F95411

A;Status: preliminary

A;Molecule type: DNA  
 A;Residues: 1-237 <KUR>  
 A;Cross-references: GB:AE006469; PIDN:AAK65856.1; PID:g14524363; GSPDB:GN00165  
 A;Experimental source: strain 1021, megaplasmid pSymA  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.;  
 Barloy-Hubler, F.; Barnett, M.J.; Becker, A.; Boistard, P.; Bothe, G.; Boutry,  
 M.; Bowser, L.; Buhrmester, J.; Cadieu, E.; Capela, D.; Chain, P.; Cowie, A.;  
 Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; Gloux, S.; Godrie, T.;  
 Goffeau, A.; Golding, B.; Gouzy, J.; Gurjal, M.; Hernandez-Lucas, I.; Hong, A.;  
 Huizar, L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.;  
 Lelaure, V.; Masuy, D.; Palm, C.; Peck, M.C.; Pohl, T.M.; Portetelle, D.;  
 Purnelle, B.; Ramsperger, U.; Surzycki, R.; Thebault, P.; Vandenbol, M.;  
 Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.; Batut, J.  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: Sma2221  
 A;Genome: plasmid

Query Match 73.9%; Score 34; DB 2; Length 237;  
 Best Local Similarity 75.0%; Pred. No. 43;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 || || ||  
 Db 76 CTAITTSC 83

RESULT 4  
 B86941  
 hypothetical protein [imported] - Mycobacterium leprae  
 C;Species: Mycobacterium leprae  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C;Accession: B86941  
 R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler,  
 P.R.; Honore, N.; Ganier, T.; Churcher, C.; Harris, D.; Mungall, K.; Basham, D.;  
 Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.M.; Devlin, K.; Duthoy, S.;  
 Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;  
 Lacroix, C.; Maclean, J.; Moule, S.; Murphy, L.; Oliver, K.; Quail, M.A.;  
 Rajandream, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.;  
 Squares, R.; Squares, S.; Stevens, K.; Taylor, K.; Whitehead, S.; Woodward,  
 J.R.; Barrell, B.G.  
 A;Title: Massive gene decay in the leprosy bacillus.  
 A;Reference number: A86909; MUID:21128732; PMID:11234002  
 A;Accession: B86941  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-317 <STO>  
 A;Cross-references: GB:AL450380; NID:g13092597; PIDN:CAC29766.1; GSPDB:GN00147  
 C;Genetics:  
 A;Gene: ML0258



Query Match 73.9%; Score 34; DB 2; Length 317;  
Best Local Similarity 62.5%; Pred. No. 55;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | | |  
Db 169 CVRLTERC 176

RESULT 5

E70623

hypothetical protein Rv1026 - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C;Accession: E70623

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;  
Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaiia, F.; Badcock, K.;  
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;  
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,  
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;  
Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete  
genome sequence.

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: E70623

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-319 <COL>

A;Cross-references: GB:Z92539; GB:AL123456; NID:g3261714; PIDN:CAB06853.1;

PID:e304624; PID:g1870002

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: Rv1026

Query Match 73.9%; Score 34; DB 2; Length 319;  
Best Local Similarity 62.5%; Pred. No. 55;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | | |  
Db 171 CVRLTERC 178

RESULT 6

T12117

polyprotein - fava bean dsRNA replicon

C;Species: Vicia faba (fava bean)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

C;Accession: T12117

R;Pfeiffer, P.

J. Gen. Virol. 79, 2349-2358, 1998

A;Title: Nucleotide sequence, genetic organization and expression strategy of  
the double-stranded RNA associated with the '447' cytoplasmic male sterility in  
Vicia faba.

A;Reference number: Z17424; MUID:98451319; PMID:9780039  
A;Accession: T12117  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-5825 <PFE>  
A;Cross-references: EMBL:AJ000929; NID:g3184155; PIDN:CAA04392.1; PID:g3184156  
A;Experimental source: virion; cultivar 447  
C;Comment: This gene product may be cleaved into several proteins including  
helicase and RNA-directed RNA polymerase.  
C;Genetics:  
A;Genome: dsRNA replicon  
C;Superfamily: fava bean dsRNA replicon polyprotein

Query Match 71.7%; Score 33; DB 2; Length 5825;  
Best Local Similarity 62.5%; Pred. No. 9.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| :|||  
Db 1909 CTCVTEKC 1916

#### RESULT 7

D37057

epithelial cell glycoprotein IIIa - guinea pig (fragment)

C;Species: Cavia porcellus (guinea pig)

C;Date: 15-Feb-1991 #sequence\_revision 15-Feb-1991 #text\_change 23-Jul-1999

C;Accession: D37057

R;Sheppard, D.; Rozzo, C.; Starr, L.; Quaranta, V.; Erle, D.J.; Pytela, R.  
J. Biol. Chem. 265, 11502-11507, 1990

A;Title: Complete amino acid sequence of a novel integrin beta subunit (beta6)  
identified in epithelial cells using the polymerase chain reaction.

A;Reference number: A37057; MUID:90307659; PMID:2365683

A;Accession: D37057

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-92 <SHE>

A;Cross-references: GB:J05522

C;Superfamily: integrin beta chain; laminin-type EGF-like homology

C;Keywords: glycoprotein

Query Match 69.6%; Score 32; DB 2; Length 92;  
Best Local Similarity 50.0%; Pred. No. 47;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| :|::|  
Db 55 CTTLTDTC 62

#### RESULT 8

S77232

hypothetical protein sll1348 - Synechocystis sp. (strain PCC 6803)

C;Species: Synechocystis sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C;Accession: S77232

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S77232

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-289 <KAN>

A;Cross-references: EMBL:D90907; GB:AB001339; NID:g1652618; PIDN:BAA17566.1; PID:d1018299; PID:g1652646

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Start codon: GTG

Query Match 69.6%; Score 32; DB 2; Length 289;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||| |:  
Db 168 CTCLEAC 175

#### RESULT 9

T39702

probable peroxisome assembly protein - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000

C;Accession: T39702

R;Wood, V.; Skelton, J.; Churcher, C.M.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, July 1999

A;Reference number: Z21870

A;Accession: T39702

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-306 <WOO>

A;Cross-references: EMBL:AL109652; PIDN:CAB51769.1; GSPDB:GN00067

A;Experimental source: strain 972h-; cosmid c17A3

C;Genetics:

A;Gene: pi037

A;Map position: 2

A;Introns: 13/1

C;Superfamily: RING finger homology

F;252-299/Domain: RING finger homology <RRN>

Query Match 69.6%; Score 32; DB 2; Length 306;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|| | |:  
Db 168 CTCLEAC 175

Db

15 CTEIDEAC 22

RESULT 10

C88940

protein C05E4.12 [imported] - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

C;Accession: C88940

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and  
www.sanger.ac.uk/Projects/C\_elegans/ for a list of authors

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103,  
1999; and Science 285, 1493, 1999

A;Accession: C88940

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-366 <STO>

A;Cross-references: GB:chr\_V; PIDN:AAB71280.1; PID:g2435573; GSPDB:GN00023;  
CESP:C05E4.12

C;Genetics:

A;Gene: C05E4.12

A;Map position: 5

Query Match	69.6%;	Score 32;	DB 2;	Length 366;
Best Local Similarity	62.5%;	Pred. No. 1.5e+02;		
Matches	5;	Conservative	2;	Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8

| ||:|:

Db 209 CFRISENC 216

RESULT 11

A70438

flagellar export protein - *Aquifex aeolicus*

C;Species: *Aquifex aeolicus*

C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C;Accession: A70438

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham,  
D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman,  
R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.

Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex*  
*aeolicus*.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: A70438

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-443 <AQF>

A;Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07494.1; PID:g2983946;  
GB:AE000657

A;Experimental source: strain VF5

C;Genetics:  
A;Gene: fliI  
C;Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase alpha chain homology  
F;192-361/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 69.6%; Score 32; DB 2; Length 443;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TRITESC 8  
||| |||  
Db 293 TRIAESC 299

#### RESULT 12

T00206

epidermis-specific protein 1 - Ciona savignyi

C;Species: Ciona savignyi

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000

C;Accession: T00206

R;Chiba, S.; Satou, Y.; Nishikata, T.; Satoh, N.

submitted to the EMBL Data Library, November 1997

A;Description: Isolation and characterization of cDNA clones for tissue-specific genes in Ciona savignyi embryos. I. epidermis-specific and muscle-specific genes.

A;Reference number: Z14123

A;Accession: T00206

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-741 <CHI>

A;Cross-references: EMBL:AB008818; PIDN:BAA23597.1

C;Superfamily: Ciona savignyi epidermis-specific protein 1; trefoil homology

F;568-610/Domain: trefoil homology <TRF>

Query Match 69.6%; Score 32; DB 2; Length 741;  
Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTRITE 6  
|||:||  
Db 237 CTRVTE 242

#### RESULT 13

B44007

aptotoxin VII - trap-door spider (Aptostichus schlingeri)

N;Alternate names: insecticidal peptide Aps VII

C;Species: Aptostichus schlingeri

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 24-May-2001

C;Accession: B44007

R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Quistad, G.B.

Toxicon 30, 1043-1050, 1992

A;Title: Identification of insecticidal peptides from venom of the trap-door spider, Aptostichus schlingeri (Ctenizidae).

A;Reference number: A44007; MUID:93069259; PMID:1440641

A;Accession: B44007

A;Molecule type: protein  
A;Residues: 1-32 <SKI>  
A;Cross-references: PIDN:AAB24048.1; PID:g259278  
A;Note: sequence extracted from NCBI backbone (NCBIP:119529)  
C;Keywords: disulfide bond; toxin; venom

Query Match 67.4%; Score 31; DB 2; Length 32;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : |  
Db 4 CARVKEAC 11

RESULT 14

JQ2199

UL50h protein - Marek's disease virus (fragment)

C;Species: Marek's disease virus

C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999

C;Accession: JQ2199

R;Yanagida, N.; Yoshida, S.; Nazerian, K.; Lee, L.F.

J. Gen. Virol. 74, 1837-1845, 1993

A;Title: Nucleotide and predicted amino acid sequences of Marek's disease virus homologues of herpes simplex virus major tegument proteins.

A;Reference number: JQ2199; MUID:93389438; PMID:8397281

A;Accession: JQ2199

A;Molecule type: DNA

A;Residues: 1-101 <YAN>

A;Cross-references: GB:L10283; NID:g388703; PIDN:AAA03146.1; PID:g388704

A;Experimental source: strain GA

Query Match 67.4%; Score 31; DB 2; Length 101;  
Best Local Similarity 50.0%; Pred. No. 78;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : |  
Db 55 CLRVTNNC 62

RESULT 15

A34398

antistasin - Mexican leech

C;Species: Haementeria officinalis (Mexican leech)

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 23-Jun-1993

C;Accession: A34398

R;Dunwiddie, C.; Thornberry, N.A.; Bull, H.G.; Sardana, M.; Friedman, P.A.; Jacobs, J.W.; Simpson, E.

J. Biol. Chem. 264, 16694-16699, 1989

A;Title: Antistasin, a leech-derived inhibitor of factor Xa. Kinetic analysis of enzyme inhibition and identification of the reactive site.

A;Reference number: A34398; MUID:89380295; PMID:2777803

A;Accession: A34398

A;Molecule type: protein

A;Residues: 1-119 <DUN>

C;Superfamily: antistasin

Query Match 67.4%; Score 31; DB 2; Length 119;  
 Best Local Similarity 50.0%; Pred. No. 89;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | : | : | |  
 Db 73 CSRLTNKC 80

Search completed: November 13, 2003, 09:53:01  
 Job time : 10.3333 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 4.58333 Seconds  
 (without alignments)  
 82.083 Million cell updates/sec

Title: US-09-228-866-9  
 Perfect score: 46  
 Sequence: 1 CTRITESC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	% Query Match	Length	DB	ID	Description
1	32	69.6	131	1	AGSR_MOUSE	P56473 mus musculu	
2	32	69.6	132	1	AGSR_HUMAN	O00253 homo sapien	
3	32	69.6	134	1	AGSR_BOVIN	P56413 bos taurus	
4	32	69.6	443	1	FLII_AQUAE	O67531 aquifex aeo	
5	32	69.6	1037	1	CAR6_HUMAN	Q9bx69 homo sapien	

6	32	69.6	1729	1	DME_ARATH	Q8lk56	arabidopsis
7	31	67.4	32	1	TXP7_APTSC	P49271	aptostichus
8	31	67.4	119	1	ANTA_HAEGH	P16242	haementeria
9	31	67.4	136	1	ANTA_HAEOF	P15358	haementeria
10	31	67.4	178	1	CHHC_BOMMO	P20730	bombyx mori
11	31	67.4	245	1	YIT8_YEAST	P40574	saccharomyc
12	31	67.4	329	1	CDK7_RAT	P51952	rattus norv
13	31	67.4	397	1	CATE_MOUSE	P70269	mus musculu
14	31	67.4	744	1	DAF4_CAEEL	P50488	caenorhabdi
15	31	67.4	760	1	YCE5_YEAST	P25574	saccharomyc
16	31	67.4	769	1	ITB2_HUMAN	P05107	homo sapien
17	31	67.4	770	1	NASE_BACSU	P42433	bacillus su
18	31	67.4	788	1	ITB6_HUMAN	P18564	homo sapien
19	31	67.4	5376	1	ZAN_MOUSE	O88799	mus musculu
20	30	65.2	120	1	Y950_AQUAE	O67084	aquifex aeo
21	30	65.2	144	1	YLX3_CAEEL	P46499	caenorhabdi
22	30	65.2	201	1	YNBA_ECOLI	P76090	escherichia
23	30	65.2	250	1	CTGL_RAT	Q9jhc6	rattus norv
24	30	65.2	281	1	T2MT_METTF	P29565	methanobact
25	30	65.2	340	1	UL20_HCMVA	P16758	human cytom
26	30	65.2	341	1	VP3_GFLV	P17768	grapevine f
27	30	65.2	349	1	DKK3_MOUSE	Q9qun9	mus musculu
28	30	65.2	350	1	DKK3_HUMAN	Q9ubp4	homo sapien
29	30	65.2	360	1	VP3_ARMV	P24820	arabis mosa
30	30	65.2	454	1	ATTY_HUMAN	P17735	homo sapien
31	30	65.2	454	1	ATTY_RAT	P04694	rattus norv
32	30	65.2	481	1	SES1_XENLA	P58003	xenopus lae
33	30	65.2	489	1	MPPB_HUMAN	O75439	homo sapien
34	30	65.2	508	1	LCK_HUMAN	P06239	homo sapien
35	30	65.2	551	1	Y900_METJA	Q58310	methanococc
36	30	65.2	565	1	FXJ2_MOUSE	Q9es18	mus musculu
37	30	65.2	635	1	SUV9_DROME	P45975	drosophila
38	30	65.2	704	1	FBL1_CHICK	O73775	gallus gall
39	30	65.2	1310	1	ACN1_HUMAN	O14525	homo sapien
40	30	65.2	1581	1	VGLP_BEV	P23052	berne virus
41	30	65.2	3579	1	STAN_DROME	Q9v5n8	drosophila
42	29	63.0	58	1	IWIT_MEDSA	P16346	medicago sa
43	29	63.0	84	1	HSPC_ELECI	P83183	eledone cir
44	29	63.0	123	1	PSCA_HUMAN	O43653	homo sapien
45	29	63.0	131	1	CHHB_BOMMO	P05688	bombyx mori

# ALIGNMENTS

## RESULT 1

### AGSR\_MOUSE

ID AGSR\_MOUSE STANDARD; PRT; 131 AA.

AC P56473; O35967;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Agouti-related protein precursor.

GN AGRP OR ART OR AGRT.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=97458244; PubMed=9311920;  
 RA Ollmann M.M., Wilson B.D., Yang Y.K., Kerns J.A., Chen Y., Gantz I.,  
 RA Barsh G.S.;  
 RT "Antagonism of central melanocortin receptors in vitro and in vivo by  
 RT agouti-related protein.";  
 RL Science 278:135-138(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97230362; PubMed=9119224;  
 RA Shutter J.R., Graham M., Kinsey A.C., Scully S., Luethy R.,  
 RA Stark K.L.;  
 RT "Hypothalamic expression of ART, a novel gene related to agouti, is  
 RT up-regulated in obese and diabetic mutant mice.";  
 RL Genes Dev. 11:593-602(1997).  
 CC -!- FUNCTION: PLAYS A ROLE IN WEIGHT HOMEOSTASIS. MAY PLAY A ROLE IN  
 CC THE REGULATION OF MELANOCORTIN RECEPTORS WITHIN THE HYPOTHALAMUS  
 CC AND ADRENAL GLAND, AND THEREFORE IN THE CENTRAL CONTROL OF  
 CC FEEDING.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ARCUATE NUCLEUS AND MEDIAN  
 CC EMINENCE, ADRENAL GLAND (MEDULLA), HYPOTHALAMUS, TESTIS, AND LUNG.  
 CC -!- INDUCTION: HYPOTHALAMIC EXPRESSION IS ELEVATED CIRCA 10 FOLD IN  
 CC OB/OB AND DB/DB MICE.  
 CC -!- SIMILARITY: BELONGS TO THE AGOUTI FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U89484; AAB68620.1; -.  
 DR EMBL; U89486; AAB68622.1; -.  
 DR MGD; MGI:892013; Agrp.  
 DR GO; GO:0005184; F:neuropeptide hormone activity; IDA.  
 DR GO; GO:0007218; P:neuropeptide signaling pathway; IDA.  
 DR GO; GO:0007582; P:physiological processes; IDA.  
 DR Pfam; PF05039; agouti; 1.  
 KW Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 131 AGOUTI-RELATED PROTEIN.  
 FT DOMAIN 86 128 CYS-RICH.  
 FT DISULFID 86 101 BY SIMILARITY.  
 FT DISULFID 93 107 BY SIMILARITY.  
 FT DISULFID 100 118 BY SIMILARITY.  
 FT DISULFID 104 128 BY SIMILARITY.  
 FT DISULFID 109 116 BY SIMILARITY.  
 SQ SEQUENCE 131 AA; 14432 MW; 25D9766D074C6834 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 131;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy            1 CTRITESC 8  
             | | : | | |  
Db            86 CVRLHESC 93

RESULT 2

AGSR\_HUMAN

ID    AGSR\_HUMAN            STANDARD;            PRT;    132 AA.  
AC    O00253; O15459;  
DT    01-NOV-1997 (Rel. 35, Created)  
DT    01-NOV-1997 (Rel. 35, Last sequence update)  
DT    15-SEP-2003 (Rel. 42, Last annotation update)  
DE    Agouti-related protein precursor.  
GN    AGRP OR ART OR AGRT.  
OS    Homo sapiens (Human).  
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX    NCBI\_TaxID=9606;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RX    MEDLINE=97230362; PubMed=9119224;  
RA    Shutter J.R., Graham M., Kinsey A.C., Scully S., Luethy R.,  
RA    Stark K.L.;  
RT    "Hypothalamic expression of ART, a novel gene related to agouti, is  
RT    up-regulated in obese and diabetic mutant mice.";  
RL    Genes Dev. 11:593-602(1997).  
RN    [2]  
RP    SEQUENCE FROM N.A.  
RC    TISSUE=Adrenal gland;  
RX    MEDLINE=97458244; PubMed=9311920;  
RA    Ollmann M.M., Wilson B.D., Yang Y.K., Kerns J.A., Chen Y., Gantz I.,  
RA    Barsh G.S.;  
RT    "Antagonism of central melanocortin receptors in vitro and in vivo by  
RT    agouti-related protein.";  
RL    Science 278:135-138(1997).  
RN    [3]  
RP    SEQUENCE FROM N.A., AND VARIANT THR-67.  
RX    MEDLINE=21488347; PubMed=11602360;  
RA    Brown A.M., Mayfield D.K., Volaufova J., Argyropoulos G.;  
RT    "The gene structure and minimal promoter of the human agouti related  
RT    protein.";  
RL    Gene 277:231-238(2001).  
RN    [4]  
RP    SEQUENCE FROM N.A.  
RA    Vink T.;  
RT    "Association between an AGRP gene polymorphism and Anorexia  
RT    Nervosa.";  
RL    Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
RN    [5]  
RP    DISULFIDE BONDS.  
RX    MEDLINE=98393470; PubMed=9724530;  
RA    Bures E.J., Hui J.O., Young Y., Chow D.T., Katta V., Rohde M.F.,  
RA    Zeni L., Rosenfeld R.D., Stark K.L., Haniu M.;  
RT    "Determination of disulfide structure in agouti-related protein (AGRP)  
RT    by stepwise reduction and alkylation.";  
RL    Biochemistry 37:12172-12177(1998).

RN [6]  
 RP STRUCTURE BY NMR OF 87-132.  
 RX MEDLINE=99297561; PubMed=10371151;  
 RA Bolin K.A., Anderson D.J., Trulson J.A., Thompson D.A., Wilken J.,  
 RA Kent S.B.H., Gantz I., Millhauser G.L.;  
 RT "NMR structure of a minimized human agouti related protein prepared  
 RT by total chemical synthesis.";  
 RL FEBS Lett. 451:125-131(1999).  
 RN [7]  
 RP STRUCTURE BY NMR OF 87-120.  
 RX MEDLINE=22052396; PubMed=12056887;  
 RA Jackson P.J., McNulty J.C., Yang Y.K., Thompson D.A., Chai B.,  
 RA Gantz I., Barsh G.S., Millhauser G.L.;  
 RT "Design, pharmacology, and NMR structure of a minimized cystine knot  
 RT with agouti-related protein activity.";  
 RL Biochemistry 41:7565-7572(2002).  
 RN [8]  
 RP VARIANT THR-67.  
 RX MEDLINE=22202398; PubMed=12213871;  
 RA Argyropoulos G., Rankinen T., Neufeld D.R., Rice T., Province M.A.,  
 RA Leon A.S., Skinner J.S., Wilmore J.H., Rao D.C., Bouchard C.;  
 RT "A polymorphism in the human agouti-related protein is associated with  
 RT late-onset obesity.";  
 RL J. Clin. Endocrinol. Metab. 87:4198-4202(2002).  
 CC -!- FUNCTION: PLAYS A ROLE IN WEIGHT HOMEOSTASIS. MAY PLAY A ROLE IN  
 CC THE REGULATION OF MELANOCORTIN RECEPTORS WITHIN THE HYPOTHALAMUS  
 CC AND ADRENAL GLAND, AND THEREFORE IN THE CENTRAL CONTROL OF  
 CC FEEDING.  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN THE ADRENAL GLAND,  
 CC SUBTHALAMIC NUCLEUS, AND HYPOTHALAMUS, WITH A LOWER LEVEL OF  
 CC EXPRESSION OCCURRING IN TESTIS, LUNG, AND KIDNEY.  
 CC -!- DISEASE: Defects in AGRP may be a cause of autosomal dominant  
 CC obesity [MIM:601665].  
 CC -!- SIMILARITY: BELONGS TO THE AGOUTI FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U88063; AAB52240.1; -.  
 DR EMBL; U89485; AAB68621.1; -.  
 DR EMBL; AF314194; AAL09457.1; -.  
 DR EMBL; AF281309; AAK96256.1; -.  
 DR PDB; 1HYK; 07-FEB-01.  
 DR PDB; 1MR0; 02-OCT-02.  
 DR Genew; HGNC:330; AGRP.  
 DR MIM; 602311; -.  
 DR MIM; 601665; -.  
 DR GO; GO:0005184; F:neuropeptide hormone activity; TAS.  
 DR GO; GO:0005102; F:receptor binding activity; TAS.  
 DR GO; GO:0007631; P:feeding behavior; TAS.  
 DR GO; GO:0007218; P:neuropeptide signaling pathway; TAS.

DR Pfam; PF05039; agouti; 1.  
 KW Signal; Disease mutation; Obesity; 3D-structure.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 132 AGOUTI-RELATED PROTEIN.  
 FT DOMAIN 87 129 CYS-RICH.  
 FT DISULFID 87 102  
 FT DISULFID 94 108  
 FT DISULFID 101 119  
 FT DISULFID 105 129  
 FT DISULFID 110 117  
 FT VARIANT 67 67 A -> T (in obesity; late onset).  
 FT /FTid=VAR\_015385.  
 FT CONFLICT 6 6 V -> L (IN REF. 2).  
 SQ SEQUENCE 132 AA; 14440 MW; 1CCBE112C3EB10F5 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 132;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | | |  
 Db 87 CVRLHESC 94

### RESULT 3

#### AGSR\_BOVIN

ID AGSR\_BOVIN STANDARD; PRT; 134 AA.  
 AC P56413;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Agouti-related protein precursor.  
 GN AGRP OR ART OR AGRT.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Oulmouden A., Petit J.M., Julien R.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: PLAYS A ROLE IN WEIGHT HOMEOSTASIS. MAY PLAY A ROLE IN  
 CC THE REGULATION OF MELANOCORTIN RECEPTORS WITHIN THE HYPOTHALAMUS  
 CC AND ADRENAL GLAND, AND THEREFORE IN THE CENTRAL CONTROL OF FEEDING  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE AGOUTI FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AJ002025; CAA05148.1; -.  
 DR Pfam; PF05039; agouti; 1.  
 KW Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 134 AGOUTI-RELATED PROTEIN.  
 FT DOMAIN 89 131 CYS-RICH.  
 FT DISULFID 89 104 BY SIMILARITY.  
 FT DISULFID 96 110 BY SIMILARITY.  
 FT DISULFID 103 121 BY SIMILARITY.  
 FT DISULFID 107 131 BY SIMILARITY.  
 FT DISULFID 112 119 BY SIMILARITY.  
 SQ SEQUENCE 134 AA; 14706 MW; F4B7AE1458B6A24B CRC64;

Query Match 69.6%; Score 32; DB 1; Length 134;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | | |  
 Db 89 CVRLHESC 96

#### RESULT 4

##### FLII\_AQUAE

ID FLII\_AQUAE STANDARD; PRT; 443 AA.  
 AC O67531;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Flagellum-specific ATP synthase (EC 3.6.3.14).  
 GN FLII OR AQ\_1595.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus."  
 RL Nature 392:353-358(1998).  
 CC -!- FUNCTION: PROBABLE CATALYTIC SUBUNIT OF A PROTEIN TRANSLOCASE FOR  
 CC FLAGELLUM-SPECIFIC EXPORT, OR A PROTON TRANSLOCASE INVOLVED IN  
 CC LOCAL CIRCUITS AT THE FLAGELLUM (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
 CC H(+) (Out).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000747; AAC07494.1; -.  
 DR PIR; A70438; A70438.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR000194; ATPase\_a/bcentre.  
 DR InterPro; IPR004100; ATPase\_a/bN.  
 DR InterPro; IPR005714; FliI\_YscN.  
 DR Pfam; PF00006; ATP-synt\_ab; 1.  
 DR Pfam; PF02874; ATP-synt\_ab\_N; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR01026; fliI\_yscN; 1.  
 DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA; 1.  
 KW Hydrolase; Hydrogen ion transport; ATP synthesis; ATP-binding;  
 KW Transport; Protein transport; Flagella; Complete proteome.  
 FT NP\_BIND 169 176 ATP (POTENTIAL).  
 SQ SEQUENCE 443 AA; 48506 MW; 62D7F5BD7C6B14A0 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 443;  
 Best Local Similarity 85.7%; Pred. No. 43;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TRITESC 8  
 ||| |||  
 Db 293 TRIAESC 299

# RESULT 5

## CAR6\_HUMAN

ID CAR6\_HUMAN STANDARD; PRT; 1037 AA.  
 AC Q9BX69;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Caspase recruitment domain protein 6.  
 GN CARD6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bertin J.;  
 RT "CARD6: a novel caspase recruitment domain (CARD) protein that  
 RT regulates apoptosis.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: May be involved in apoptosis.  
 CC -!- SIMILARITY: Contains 1 CARD domain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; AF356193; AAK32718.1; -.
DR Genew; HGNC:16394; CARD6.
DR InterPro; IPR001315; CARD.
DR Pfam; PF00619; CARD; 1.
DR SMART; SM00114; CARD; 1.
DR PROSITE; PS50209; CARD; 1.
KW Apoptosis.
FT DOMAIN      3      94      CARD.
FT DOMAIN      201     281     ASP/GLU-RICH.
SQ SEQUENCE    1037 AA; 116493 MW; 592C189CA51EA90F CRC64;

```

```

Query Match          69.6%; Score 32; DB 1; Length 1037;
Best Local Similarity 83.3%; Pred. NO. 1e+02;
Matches      5; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CTRITE 6
        |||:||
Db      872 CTRVTE 877

```

# RESULT 6

## DME\_ARATH

```

ID DME_ARATH STANDARD; PRT; 1729 AA.
AC Q8LK56; Q9LZ67; Q9LZ68; Q9LZ69;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcriptional activator DEMETER (DNA glycosylase-related protein
DE DME).
GN DME OR AT5G04560/AT5G04570/AT5G04580 OR
GN T32M21.160/T32M21.170/T32M21.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND CHARACTERIZATION.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=22145911; PubMed=12150995;
RA Choi Y., Gehring M., Johnson L., Hannon M., Harada J.J.,
RA Goldberg R.B., Jacobsen S.E., Fischer R.L.;
RT "DEMETER, a DNA glycosylase domain protein, is required for endosperm
RT gene imprinting and seed viability in Arabidopsis.";
RL Cell 110:33-42(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,

```

RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,  
 RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,  
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
 RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,  
 RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,  
 RA Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,  
 RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;  
 RT "Sequence and analysis of chromosome 5 of the plant *Arabidopsis*  
 RT *thaliana*.";  
 RL Nature 408:823-826(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=cv. Columbia;  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RT "Arabidopsis thaliana full-length cDNA.";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Transcriptional activator involved in gene imprinting.  
 CC Allows the expression of the maternal copy of the imprinted MEA  
 CC gene before fertilization, possibly by antagonizing or suppressing  
 CC DNA methylation on target promoter. Probably acts by nicking the  
 CC MEA promoter. Required for stable reproducible patterns of floral  
 CC and vegetative development.  
 CC -!- COFACTOR: Binds a 4Fe-4S cluster which is probably involved in the  
 CC proper positioning of the protein along the DNA strand (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q8LK56-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q8LK56-2; Sequence=VSP\_007455;  
 CC Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Mainly expressed in immature flower buds, then  
 CC decreases as the flower matures. Expressed in the ovule carpels,  
 CC but not expressed in pollen stamens. Expressed in developing and  
 CC mature ovules (stages 12-14), then strongly decreases after  
 CC fertilization.  
 CC -!- DEVELOPMENTAL STAGE: Maternally expressed. Expressed primarily in  
 CC the central cell of gametophyte before fertilization. Not  
 CC expressed in endosperm and embryo after fertilization.  
 CC -!- DOMAIN: The DEMETER domain, which is present in proteins of the  
 CC subfamily, is related to the J-domain, but lacks some important  
 CC conserved residues.  
 CC -!- MISCELLANEOUS: Although strongly related to DNA glycosylase



CC proteins, it differs from these proteins because of its large size  
 CC and its unique N-terminal basic domain. The DNA repair function  
 CC has not been proved and may not exist.  
 CC -!- SIMILARITY: BELONGS TO THE DNA GLYCOSYLASE FAMILY. DEMETER  
 CC SUBFAMILY.  
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF521596; AAM77215.1; -.  
 DR EMBL; AL162875; CAB85562.1; ALT\_SEQ.  
 DR EMBL; AL162875; CAB85563.1; ALT\_SEQ.  
 DR EMBL; AL162875; CAB85564.1; ALT\_SEQ.  
 DR EMBL; AK117994; BAC42629.1; -.  
 DR InterPro; IPR003265; Endo\_3c.  
 DR InterPro; IPR003651; FeS\_bind.  
 DR Pfam; PF00730; HhH-GPD; 1.  
 DR SMART; SM00478; ENDO3c; 1.  
 DR SMART; SM00525; FES; 1.  
 DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; FALSE\_NEG.  
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW 4Fe-4S; Iron-sulfur; Alternative splicing.  
 FT DOMAIN 697 796 DEMETER.  
 FT DOMAIN 33 108 LYS-RICH (BASIC).  
 FT DOMAIN 215 367 GLN-RICH.  
 FT METAL 1371 1371 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 1378 1378 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 1381 1381 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 1387 1387 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT VARSPLIC 1 1313 Missing (in isoform 2).  
 FT /FTid=VSP\_007455.  
 FT CONFLICT 1421 1421 F -> Y (IN REF. 3).  
 SQ SEQUENCE 1729 AA; 192888 MW; AD9D7A91FDB4E251 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1729;  
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTRITES 7  
 || ||||  
 Db 1464 CTEITES 1470

RESULT 7  
 TXP7\_APTSC  
 ID TXP7\_APTSC STANDARD; PRT; 32 AA.  
 AC P49271;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Aptotoxin VII (Paralytic peptide VII) (PP VII).  
 OS Aptostichus schlingeri (Trap-door spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Mygalomorphae; Cyrtaucheniidae; Apomastus.  
 OX NCBI\_TaxID=12944;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RX MEDLINE=93069259; PubMed=1440641;  
 RA Skinner W.S., Dennis P.A., Li J.P., Quistad G.B.;  
 RT "Identification of insecticidal peptides from venom of the trap-door  
 RT spider, Aptostichus schlingeri (Ctenizidae).";  
 RL Toxicon 30:1043-1050(1992).  
 CC -!- FUNCTION: IS BOTH PARALYTIC AND LETHAL, WHEN INJECTED INTO  
 CC LEPIDOPTERAN LARVAE. IS A SLOWER ACTING TOXIN, BEING LETHAL AT 24  
 CC HR, BUT NOT PARALYTIC AT 1 HR POST-INJECTION.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- PTM: THREE DISULFIDE BONDS ARE PRESENT.  
 CC -!- MISCELLANEOUS: LD(50) IS 1.40 MG/KG BY SUBCUTANEOUS INJECTION.  
 CC -!- SIMILARITY: TO APTOTOXIN III.  
 DR PIR; B44007; B44007.  
 KW Toxin; Neurotoxin.  
 SQ SEQUENCE 32 AA; 3537 MW; 2AFB15230F06BCF6 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 32;  
 Best Local Similarity 50.0%; Pred. No. 4.9;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : |  
 Db 4 CARVKEAC 11

#### RESULT 8

ANTA\_HAEGH

ID ANTA\_HAEGH STANDARD; PRT; 119 AA.  
 AC P16242;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ghilanten.  
 OS Haementeria ghilianii (Amazon leech).  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;  
 OC Rhynchobdellida; Glossiphoniidae; Haementeria.  
 OX NCBI\_TaxID=6409;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Saliva;  
 RX MEDLINE=90165947; PubMed=2306252;  
 RA Blankenship D.T., Brankamp R.G., Manley G.D., Cardin A.D.;  
 RT "Amino acid sequence of ghilanten: anticoagulant-antimetastatic  
 RT principle of the South American leech, Haementeria ghilianii.";  
 RL Biochem. Biophys. Res. Commun. 166:1384-1389(1990).  
 CC -!- FUNCTION: THIS HIGHLY DISULFIDE-BONDED PROTEIN IS A POTENT  
 CC INHIBITOR OF FACTOR XA. MAY HAVE THERAPEUTIC UTILITY AS AN  
 CC ANTICOAGULANT. ALSO EXHIBITS A STRONG METASTATIC ACTIVITY.

CC -!- MISCELLANEOUS: BINDS TO HEPARIN-AGAROSE, BINDS TO SULFATED  
 CC GLYCOCONJUGATES.  
 CC -!- SIMILARITY: BELONGS TO THE ANTISTASIN FAMILY.  
 DR PIR; A34816; A34816.  
 DR HSSP; P15358; 1SKZ.  
 DR InterPro; IPR004094; Antistasin.  
 DR Pfam; PF02822; Antistasin; 2.  
 KW Serine protease inhibitor; Repeat; Heparin-binding;  
 KW Pyrrolidone carboxylic acid.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT DOMAIN 2 110 2 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 2 55 1.  
 FT REPEAT 56 110 2.  
 FT DOMAIN 97 100 HEPARIN-BINDING (POTENTIAL).  
 FT DOMAIN 111 118 HEPARIN-BINDING (POTENTIAL).  
 FT ACT\_SITE 34 35 REACTIVE BOND (BY SIMILARITY).  
 FT ACT\_SITE 89 90 REACTIVE BOND (BY SIMILARITY).  
 FT DISULFID 8 19 BY SIMILARITY.  
 FT DISULFID 13 26 BY SIMILARITY.  
 FT DISULFID 28 48 BY SIMILARITY.  
 FT DISULFID 33 51 BY SIMILARITY.  
 FT DISULFID 37 53 BY SIMILARITY.  
 FT DISULFID 62 73 BY SIMILARITY.  
 FT DISULFID 67 80 BY SIMILARITY.  
 FT DISULFID 82 103 BY SIMILARITY.  
 FT DISULFID 88 106 BY SIMILARITY.  
 FT DISULFID 92 108 BY SIMILARITY.  
 SQ SEQUENCE 119 AA; 13317 MW; 5A94805DBBB850EF CRC64;

Query Match 67.4%; Score 31; DB 1; Length 119;  
 Best Local Similarity 50.0%; Pred. No. 18;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | : | : | |  
 Db 73 CSRLTNKC 80

# RESULT 9

## ANTA\_HAEOF

ID ANTA\_HAEOF STANDARD; PRT; 136 AA.  
 AC P15358; Q9TWQ8; Q9TX45;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Antistasin precursor (ATS) (Blood coagulation factor Xa/proclotting  
 DE enzyme inhibitor).  
 OS Haementeria officinalis (Mexican leech).  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;  
 OC Rhynchobdellida; Glossiphoniidae; Haementeria.  
 OX NCBI\_TaxID=6410;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89252921; PubMed=2470652;  
 RA Han J.H., Law S.W., Keller P.M., Kniskern P.J., Silberklang M.,  
 RA Tung J.S., Gasic T.B., Gasic G.J., Friedman P.A., Ellis R.W.;  
 RT "Cloning and expression of cDNA encoding antistasin, a leech-derived

RT protein having anti-coagulant and anti-metastatic properties.";  
 RL Gene 75:47-57(1989).  
 RN [2]  
 RP SEQUENCE OF 18-136.  
 RC TISSUE=Saliva;  
 RX MEDLINE=88273105; PubMed=3164720;  
 RA Nutt E., Gasic T., Rodkey J., Gasic G.J., Jacobs J.W., Friedman P.A.,  
 RA Simpson E.;  
 RT "The amino acid sequence of antistasin. A potent inhibitor of factor  
 RT Xa reveals a repeated internal structure.";  
 RL J. Biol. Chem. 263:10162-10167(1988).  
 RN [3]  
 RP SEQUENCE OF 18-136.  
 RC TISSUE=Saliva;  
 RX MEDLINE=94097222; PubMed=8271959;  
 RA Dunwiddie C.T., Waxman L., Vlasuk G.P., Friedman P.A.;  
 RT "Purification and characterization of inhibitors of blood coagulation  
 RT factor Xa from hematophagous organisms.";  
 RL Meth. Enzymol. 223:291-312(1993).  
 RN [4]  
 RP REACTIVE SITE.  
 RX MEDLINE=89380295; PubMed=2777803;  
 RA Dunwiddie C., Thornberry N.A., Bull H.G., Sardana M., Friedman P.A.,  
 RA Jacobs J.W., Simpson E.;  
 RT "Antistasin, a leech-derived inhibitor of factor Xa. Kinetic analysis  
 RT of enzyme inhibition and identification of the reactive site.";  
 RL J. Biol. Chem. 264:16694-16699(1989).  
 RN [5]  
 RP SULFATIDE-BINDING.  
 RX MEDLINE=89308627; PubMed=2745433;  
 RA Holt G.D., Krivan H.C., Gasic G.J., Ginsburg V.;  
 RT "Antistasin, an inhibitor of coagulation and metastasis, binds to  
 RT sulfatide (Gal(3-SO4) beta 1-1Cer) and has a sequence homology with  
 RT other proteins that bind sulfated glycoconjugates.";  
 RL J. Biol. Chem. 264:12138-12140(1989).  
 RN [6]  
 RP MUTAGENESIS.  
 RX MEDLINE=93075053; PubMed=1445252;  
 RA Hofmann K.J., Nutt E.M., Dunwiddie C.;  
 RT "Site-directed mutagenesis of the leech-derived factor Xa inhibitor  
 RT antistasin. Probing of the reactive site.";  
 RL Biochem. J. 287:943-949(1992).  
 RN [7]  
 RP MUTAGENESIS.  
 RX MEDLINE=94353372; PubMed=8073407;  
 RA Theunissen H.J., Dijkema R., Swinkels J.C., de Poorter T.L.,  
 RA Vink P.M., van Dinther T.G.;  
 RT "Mutational analysis of antistasin, an inhibitor of blood coagulation  
 RT factor Xa derived from the Mexican leech *Haementeria officinalis*.";  
 RL Thromb. Res. 75:41-50(1994).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 24-127.  
 RX MEDLINE=97459903; PubMed=9311976;  
 RA Lapatto R., Krengel U., Schreuder H.A., Arkema A., de Boer B.,  
 RA Kalk K.H., Hol W.G.J., Grootenhuys P.D.J., Mulders J.W.M., Dijkema R.,  
 RA Theunissen H.J.M., Dijkstra B.W.;  
 RT "X-ray structure of antistasin at 1.9-A resolution and its modelled

```

RT    complex with blood coagulation factor Xa.";
RL    EMBO J. 16:5151-5161(1997).
CC    -!- FUNCTION: THIS HIGHLY DISULFIDE-BONDED PROTEIN IS A POTENT
CC          INHIBITOR OF FACTOR XA. MAY HAVE THERAPEUTIC UTILITY AS AN
CC          ANTICOAGULANT. ALSO EXHIBITS A STRONG METASTATIC ACTIVITY.
CC    -!- MISCELLANEOUS: BINDS TO HEPARIN-AGAROSE, BINDS TO SULFATED
CC          GLYCOCONJUGATES.
CC    -!- MISCELLANEOUS: AT LEAST FOUR ISOFORMS OF ANTISTASIN HAVE BEEN
CC          IDENTIFIED IN LEECH SALIVARY GLAND EXTRACTS, WHICH DIFFER BY 1 OR
CC          2 AA RESIDUES.
CC    -!- SIMILARITY: BELONGS TO THE ANTISTASIN FAMILY.
CC    -----
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----
DR    EMBL; M24422; AAA29192.1; -.
DR    EMBL; M24423; AAA29193.1; -.
DR    PIR; A28806; A28806.
DR    PDB; 1SKZ; 22-OCT-97.
DR    InterPro; IPR004094; Antistasin.
DR    Pfam; PF02822; Antistasin; 2.
KW    Serine protease inhibitor; Repeat; Heparin-binding; Blood coagulation;
KW    Signal; 3D-structure; Pyrrolidone carboxylic acid.
FT    SIGNAL          1      17
FT    CHAIN          18     136      ANTISTASIN.
FT    MOD_RES        18      18      PYRROLIDONE CARBOXYLIC ACID.
FT    DOMAIN         19     127      2 X APPROXIMATE TANDEM REPEATS.
FT    REPEAT         19      72      1.
FT    REPEAT         73     127      2.
FT    ACT_SITE       51      52      REACTIVE BOND.
FT    ACT_SITE      106     107      REACTIVE BOND.
FT    DOMAIN        114     117      HEPARIN-BINDING (POTENTIAL).
FT    DOMAIN        128     135      HEPARIN-BINDING (POTENTIAL).
FT    DISULFID       25      36
FT    DISULFID       30      43
FT    DISULFID       45      65
FT    DISULFID       50      68
FT    DISULFID       54      70
FT    DISULFID       79      90
FT    DISULFID       84      97
FT    DISULFID       99     120
FT    DISULFID      105     123
FT    DISULFID      109     125
FT    VARIANT        22      22      G -> R (IN ISOFORM B).
FT    VARIANT        47      47      G -> E.
FT    VARIANT        52      52      M -> V.
FT    VARIANT        71      71      R -> I.
FT    HELIX          25      28
FT    TURN           32      33
FT    STRAND         35      35
FT    TURN           38      40
FT    STRAND         45      45

```

FT	TURN	55	56
FT	STRAND	58	60
FT	TURN	62	63
FT	STRAND	66	70
FT	HELIX	81	83
FT	TURN	86	87
FT	STRAND	88	89
FT	TURN	92	94
FT	STRAND	99	100
FT	STRAND	113	115
FT	TURN	117	118
FT	STRAND	121	125
SQ	SEQUENCE	136 AA;	15225 MW; 582AF009ED9A0291 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 136;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | : | : | |  
 Db 90 CSRLTNKC 97

# RESULT 10

CHHC\_BOMMO

ID CHHC\_BOMMO STANDARD; PRT; 178 AA.

AC P20730;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chorion class high-cysteine HCB protein 13 precursor (HC-B.13).

OS Bombyx mori (Silk moth).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Bombycidae; Bombyx.

OX NCBI\_TaxID=7091;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85083111; PubMed=6439880;

RA Rodakis G.C., Lecanidou R., Eickbush T.H.;

RT "Diversity in a chorion multigene family created by tandem

RT duplications and a putative gene-conversion event.";

RL J. Mol. Evol. 20:265-273(1984).

CC -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE  
 CC SILK MOTH.

CC -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH  
 CC BELONG CLASSES B, CB AND HCB.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X01068; -; NOT\_ANNOTATED\_CDS.

DR PIR; A23219; A23219.  
 DR HSSP; P01180; 1NPO.  
 DR InterPro; IPR002635; Chorion.  
 DR Pfam; PF01723; Chorion; 1.  
 KW Eggshell; Chorion; Repeat; Multigene family; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 178 CHORION CLASS HIGH-CYSTEINE HCB PROTEIN  
 FT 13.  
 FT DOMAIN 22 46 LEFT ARM.  
 FT DOMAIN 47 110 CENTRAL DOMAIN.  
 FT DOMAIN 111 178 RIGHT ARM (GLY-RICH TANDEM REPEATS).  
 SQ SEQUENCE 178 AA; 16077 MW; 8AF703E0F65D3096 CRC64;  
  
 Query Match 67.4%; Score 31; DB 1; Length 178;  
 Best Local Similarity 62.5%; Pred. No. 27;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
 Qy 1 CTRITESC 8  
 | ||:|  
 Db 104 CVGITQSC 111

RESULT 11

YIT8\_YEAST

ID YIT8\_YEAST STANDARD; PRT; 245 AA.  
 AC P40574;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 28.4 kDa protein in MET28-STAl intergenic region.  
 GN YIR018W.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FL100;  
 RX MEDLINE=96221312; PubMed=8665859;  
 RA Kuras L., Cherest H., Surdin-Kerjan Y., Thomas D.;  
 RT "A heteromeric complex containing the centromere binding factor 1 and  
 RT two basic leucine zipper factors, Met4 and Met28, mediates the  
 RT transcription activation of yeast sulfur metabolism.";  
 RL EMBO J. 15:2519-2529(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX PubMed=9169870;  
 RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,  
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,  
 RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,  
 RA Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,  
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX.";  
 RL Nature 387:84-87(1997).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Belongs to the bZIP family.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U17015; AAC49426.1; -.
DR EMBL; Z37996; CAA86089.1; -.
DR PIR; S48363; S48363.
DR SGD; S0001457; YAP5.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; IDA.
DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS50217; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW Hypothetical protein; DNA-binding; Nuclear protein.
FT DNA_BIND      63      82      BASIC MOTIF.
FT CONFLICT      173     173      P -> S (IN REF. 1).
SQ SEQUENCE      245 AA;  28386 MW;  02F78E1963982E0D CRC64;

```

```

Query Match          67.4%;  Score 31;  DB 1;  Length 245;
Best Local Similarity 62.5%;  Pred. No. 37;
Matches      5;  Conservative      1;  Mismatches      2;  Indels      0;  Gaps      0;

```

```

Qy      1 CTRITESC 8
        || | :||
Db      230 CTNIDKSC 237

```

# RESULT 12

```

CDK7_RAT
ID CDK7_RAT          STANDARD;          PRT;   329 AA.
AC P51952;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein kinase 7 (EC 2.7.1.-) (CDK-activating kinase)
DE (CAK) (TFIIH basal transcription factor complex kinase subunit) (39
DE protein kinase) (P39 Mo15) (Fragment).
GN CDK7 OR CAK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RA Wu L., Hall F.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Cyclin-dependent kinases (CDKs) are activated by the
CC binding to a cyclin and mediate the progression through the cell

```



```

CC      cycle. Each different complex controls a specific transition
CC      between two subsequent phases in the cell cycle. CDK7 is the
CC      catalytic subunit of the CDK-activating kinase (CAK) complex, a
CC      serine-threonine kinase. CAK activates the cyclin-associated
CC      kinases CDC2/CDK1, CDK2, CDK4 and CDK6 by threonine
CC      phosphorylation. CAK complexed to the core-TFIID basal
CC      transcription factor activates RNA polymerase II by serine
CC      phosphorylation of the repetitive carboxyl-terminus domain (CTD)
CC      of its large subunit (POLR2A), allowing its escape from the
CC      promoter and elongation of the transcripts. Involved in cell cycle
CC      control and in RNA transcription by RNA polymerase II. Its
CC      expression and activity are constant throughout the cell cycle (By
CC      similarity).
CC      -!- ENZYME REGULATION: Phosphorylation at Thr-170 is required for
CC      enzymatic activity (By similarity).
CC      -!- SUBUNIT: Associates primarily with cyclin H and MAT1 to form the
CC      CAK complex. CAK can further associate with the core-TFIID to
CC      form the TFIID basal transcription factor (By similarity).
CC      -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC      -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC      CDC2/CDKX SUBFAMILY.
CC      -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION, A STOP CODON WAS READ
CC      THROUGH IN POSITION 313 TO MAXIMIZE SIMILARITIES WITH OTHER
CC      SPECIES CDK7.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X83579; CAA58562.1; ALT_SEQ.
DR      HSSP; P24941; 1B38.
DR      InterPro; IPR000719; Prot_kinase.
DR      InterPro; IPR002290; Ser_thr_pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      ProDom; PD000001; Prot_kinase; 1.
DR      SMART; SM00220; S_TKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW      Transferase; Serine/threonine-protein kinase; ATP-binding; Meiosis;
KW      Phosphorylation; Cell cycle; Cell division; Nuclear protein;
KW      Transcription regulation.
FT      NON_TER      1      1
FT      DOMAIN      4      287      PROTEIN KINASE.
FT      NP_BIND      10     18      ATP (BY SIMILARITY).
FT      BINDING      33     33      ATP (BY SIMILARITY).
FT      ACT_SITE     129    129      BY SIMILARITY.
FT      MOD_RES      156    156      PHOSPHORYLATION (BY SIMILARITY).
FT      MOD_RES      162    162      PHOSPHORYLATION (BY SIMILARITY).
FT      NON_TER      329    329
SQ      SEQUENCE     329 AA;  37164 MW;  BBA38FD074881B0F CRC64;

```

Query Match

67.4%; Score 31; DB 1; Length 329;

Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITES 7  
      |||||  
Db 273 CTRITAS 279

RESULT 13

CATE\_MOUSE

ID CATE\_MOUSE STANDARD; PRT; 397 AA.  
AC P70269; O35647;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cathepsin E precursor (EC 3.4.23.34).  
GN CTSE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Spleen;  
RX MEDLINE=97324100; PubMed=9180269;  
RA Tatnell P.J., Lees W.E., Kay J.;  
RT "Cloning, expression and characterisation of murine procathepsin E.";  
RL FEBS Lett. 408:62-66(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RA Tatnell P.J., Roth W., Duessing J., Kay J., Peters C.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN  
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.  
CC -!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader  
CC specificity.  
CC -!- SUBUNIT: Homodimer; disulfide-linked (By similarity).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X97399; CAA66056.1; -.  
DR EMBL; Y10928; CAA71859.1; -.  
DR HSSP; P00794; 4CMS.  
DR MEROPS; A01.010; -.  
DR MGD; MGI:107361; Ctse.  
DR InterPro; IPR001969; Aspprotease\_site.  
DR InterPro; IPR001461; AspproteaseA1.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.

DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 18 BY SIMILARITY.  
 FT PROPEP 19 59 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 60 397 CATHEPSIN E.  
 FT ACT\_SITE 97 97 BY SIMILARITY.  
 FT ACT\_SITE 282 282 BY SIMILARITY.  
 FT DISULFID 61 61 INTERCHAIN (PROBABLE).  
 FT DISULFID 110 115 BY SIMILARITY.  
 FT DISULFID 273 277 BY SIMILARITY.  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 297 297 H -> Q (IN REF. 2).  
 SQ SEQUENCE 397 AA; 42932 MW; 83993FFE3AB36105 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 397;  
 Best Local Similarity 71.4%; Pred. No. 61;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TRITESC 8  
 ||::|||  
 Db 55 TRLSESC 61

#### RESULT 14

##### DAF4\_CAEEL

ID DAF4\_CAEEL STANDARD; PRT; 744 AA.  
 AC P50488;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cell-surface receptor daf-4 precursor (EC 2.7.1.37).  
 GN DAF-4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94019813; PubMed=8413626;  
 RA Estevez M., Attisano L., Wrana J.L., Albert P.S., Massague J.,  
 RA Riddle D.L.;  
 RT "The daf-4 gene encodes a bone morphogenetic protein receptor  
 RT controlling C. elegans dauer larva development.";  
 RL Nature 365:644-649(1993).  
 CC -!- FUNCTION: INVOLVED IN TGF-BETA PATHWAY. MAY BE A RECEPTOR FOR DAF-  
 CC 7. REGULATES DAUER LARVA DEVELOPMENT.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC TGFB RECEPTOR SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L23110; AAA03544.1; -.  
 DR PIR; S38279; S38279.  
 DR InterPro; IPR000472; Activin\_rec.  
 DR InterPro; IPR000333; Actn\_receptorII.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00653; ACTIVIN2R.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Transmembrane; Glycoprotein; Signal; Developmental protein;  
 KW Alternative splicing.  
 FT SIGNAL 1 47 POTENTIAL.  
 FT CHAIN 48 744 CELL-SURFACE RECEPTOR DAF-4.  
 FT DOMAIN 48 253 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 254 274 POTENTIAL.  
 FT DOMAIN 275 744 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 306 603 PROTEIN KINASE.  
 FT NP\_BIND 312 320 ATP (BY SIMILARITY).  
 FT BINDING 338 338 ATP (BY SIMILARITY).  
 FT ACT\_SITE 440 440 BY SIMILARITY.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 744 AA; 84478 MW; 942DC28D204569AC CRC64;

Query Match 67.4%; Score 31; DB 1; Length 744;  
 Best Local Similarity 62.5%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | | | |  
 Db 585 CARITAGC 592

RESULT 15  
 YCE5\_YEAST  
 ID YCE5\_YEAST STANDARD; PRT; 760 AA.  
 AC P25574;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Hypothetical 87.2 kDa protein in APA1/DTP-PDI1 intergenic region.  
 GN YCL045C OR YCL45C OR YCL315.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=92397595; PubMed=1523890;  
 RA Scherens B., Messenguy F., Gigot D., Dubois E.;  
 RT "The complete sequence of a 9,543 bp segment on the left arm of  
 RT chromosome III reveals five open reading frames including glucokinase  
 RT and the protein disulfide isomerase."  
 RL Yeast 8:577-586(1992).  
 RN [2]  
 RP SEQUENCE OF 683-760 FROM N.A.  
 RA Grenson M., Jauniaux J.-C., Urrestarazu L.A.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: SOME, TO S.POMBE SPAC25H1.07.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59720; CAA42370.1; -.  
 DR PIR; S19374; S19374.  
 DR SGD; S0000550; YCL045C.  
 KW Hypothetical protein.  
 SQ SEQUENCE 760 AA; 87181 MW; 56F2B5A7186BDF7A CRC64;

Query Match 67.4%; Score 31; DB 1; Length 760;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITES 7  
 |||||  
 Db 709 CTRITPS 715

Search completed: November 13, 2003, 09:46:38  
 Job time : 5.58333 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 21.0833 Seconds  
 (without alignments)  
 97.917 Million cell updates/sec

Title: US-09-228-866-9  
 Perfect score: 46  
 Sequence: 1 CTRITESC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	39	84.8	975	5	Q9NKT8	Q9nkt8 leishmania
2	38	82.6	628	5	Q8SR00	Q8sr00 encephalito
3	36	78.3	416	12	Q9DPG6	Q9dpg6 avian ortho
4	36	78.3	416	12	Q8JJZ5	Q8jjz5 muscovy duc
5	36	78.3	416	12	Q9DGX2	Q9dgx2 avian ortho
6	36	78.3	416	12	Q9DPG7	Q9dpg7 avian ortho
7	36	78.3	416	12	Q9DPG8	Q9dpg8 avian ortho
8	36	78.3	416	12	O72459	O72459 avian ortho
9	36	78.3	416	12	Q9DPH0	Q9dph0 avian ortho
10	36	78.3	416	12	Q9DPH1	Q9dph1 avian ortho
11	36	78.3	416	12	O72460	O72460 avian ortho
12	36	78.3	416	12	Q9YL31	Q9yl31 avian ortho
13	36	78.3	416	12	Q9DPG9	Q9dpg9 avian ortho
14	36	78.3	416	12	Q9E6F8	Q9e6f8 avian ortho
15	36	78.3	416	12	Q9DL59	Q9dl59 avian ortho
16	36	78.3	416	12	Q9DPH2	Q9dph2 avian ortho
17	35	76.1	95	5	Q24060	Q24060 drosophila
18	35	76.1	98	5	Q24077	Q24077 drosophila
19	35	76.1	153	11	Q8VHC4	Q8vhc4 mus musculu

20	35	76.1	186	16	Q9PC85	Q9pc85 xylella fas
21	35	76.1	523	16	Q9KYG3	Q9kyg3 streptomyce
22	35	76.1	3542	5	Q9U5M2	Q9u5m2 plasmodium
23	34	73.9	114	16	Q8RGG0	Q8rgg0 fusobacteri
24	34	73.9	237	16	Q92XP6	Q92xp6 rhizobium m
25	34	73.9	317	16	O69585	O69585 mycobacteri
26	34	73.9	319	16	P96374	P96374 mycobacteri
27	34	73.9	454	5	O44021	O44021 plasmodium
28	34	73.9	602	5	Q8I4S2	Q8i4s2 plasmodium
29	34	73.9	797	13	Q8UW62	Q8uw62 oreochromis
30	33	71.7	71	13	Q90WY7	Q90wy7 coturnix co
31	33	71.7	154	13	Q9PWG2	Q9pwg2 gallus gall
32	33	71.7	165	13	Q9W7R0	Q9w7r0 gallus gall
33	33	71.7	208	16	Q8Y1Y4	Q8yly4 ralstonia s
34	33	71.7	238	5	O76510	O76510 cryptospori
35	33	71.7	270	17	Q8TIA6	Q8tia6 methanosarc
36	33	71.7	278	10	Q9SQT3	Q9sqt3 arabidopsis
37	33	71.7	297	10	Q8L8T9	Q8l8t9 arabidopsis
38	33	71.7	453	4	O14586	O14586 homo sapien
39	33	71.7	453	10	Q9LU91	Q9lu91 arabidopsis
40	33	71.7	453	10	Q8L720	Q8l720 arabidopsis
41	33	71.7	539	2	O85280	O85280 ehrlichia r
42	33	71.7	692	5	Q8SWW9	Q8sww9 drosophila
43	33	71.7	816	12	Q9E1W9	Q9elw9 cercopithec
44	33	71.7	5825	10	O82731	O82731 vicia faba
45	32	69.6	78	11	Q9QXJ3	Q9qxj3 rattus norv

# ALIGNMENTS

## RESULT 1

Q9NKT8

ID Q9NKT8 PRELIMINARY; PRT; 975 AA.

AC Q9NKT8;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE L6202.5.

GN L6202.5.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI\_TaxID=5664;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Friedlin;

RA Myler P.J.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Friedlin;

RA Worthey E.A., Sisk E., Hixson G., Kiser P., Rickel E., Hassebrock M.,

RA Cawthra J., Sunkin S., Stuart K.D., Myler P.J.;

RT "Direct Submission.";

RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AC005802; AAF31048.1; -.

DR EMBL; AC125735; AAM68996.1; -.

SQ SEQUENCE 975 AA; 101922 MW; 2C35E226868FFD49 CRC64;  
Query Match 84.8%; Score 39; DB 5; Length 975;  
Best Local Similarity 75.0%; Pred. No. 7.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
|||:| ||  
Db 304 CTRLTSSC 311

RESULT 2

Q8SR00

ID Q8SR00 PRELIMINARY; PRT; 628 AA.  
AC Q8SR00;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Dynamin-like vacuolar protein sorting protein.  
GN ECU10\_1700I.  
OS Encephalitozoon cuniculi.  
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.  
OX NCBI\_TaxID=6035;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GB-M1;  
RA Genoscope;  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GB-M1;  
RX MEDLINE=21576510; PubMed=11719806;  
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,  
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
RA Weissenbach J., Vivares C.P.;  
RT "Genome sequence and gene compaction of the eukaryote parasite  
RT Encephalitozoon cuniculi.";  
RL Nature 414:450-453(2001).  
DR EMBL; AL590449; CAD25891.1; -.  
DR InterPro; IPR001401; Dynamin.  
DR InterPro; IPR000375; Dynamin\_central.  
DR InterPro; IPR003130; GED.  
DR Pfam; PF00350; dynamin; 1.  
DR Pfam; PF01031; dynamin\_2; 1.  
DR Pfam; PF02212; GED; 1.  
DR SMART; SM00053; DYNc; 1.  
DR SMART; SM00302; GED; 1.  
DR PROSITE; PS00410; DYNAMIN; 1.  
SQ SEQUENCE 628 AA; 71166 MW; 0E451D33EF2C717A CRC64;

Query Match 82.6%; Score 38; DB 5; Length 628;  
Best Local Similarity 75.0%; Pred. No. 8.1;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
||:|||| |



Db 149 CTKITEMC 156

RESULT 3

Q9DPG6

ID Q9DPG6 PRELIMINARY; PRT; 416 AA.  
AC Q9DPG6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Sigma A.  
OS Avian orthoreovirus.  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=38170;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=601SI;  
RA Liu H.J., Huang P.H.;  
RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
reovirus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF294769; AAG44968.1; -.  
DR InterPro; IPR004317; Sigma\_1\_2.  
DR Pfam; PF03084; Sigma\_1\_2; 1.  
SQ SEQUENCE 416 AA; 46083 MW; E238CDCA86F6B8F9 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : ||  
Db 189 CARLTQSC 196

RESULT 4

Q8JJZ5

ID Q8JJZ5 PRELIMINARY; PRT; 416 AA.  
AC Q8JJZ5;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Sigma A protein.  
OS Muscovy duck reovirus.  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=77153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=89026;  
RX MEDLINE=21959063; PubMed=11961275;  
RA Kuntz-Simon G., Le Gall-Recule G., de Boisseson C., Jestin V.;  
RT "Muscovy duck reovirus sigma C protein is atypically encoded by the  
smallest genome segment.";  
RL J. Gen. Virol. 83:1189-1200(2002).  
DR EMBL; AJ278102; CAC81941.1; -.  
DR InterPro; IPR004317; Sigma\_1\_2.  
DR Pfam; PF03084; Sigma\_1\_2; 1.

SQ SEQUENCE 416 AA; 46160 MW; 06F9F80FA25555C7 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : | |  
Db 189 CARLTQSC 196

RESULT 5

Q9DGX2

ID Q9DGX2 PRELIMINARY; PRT; 416 AA.  
AC Q9DGX2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Sigma A.  
OS Avian orthoreovirus.  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=38170;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OS161, and 919;  
RA Liu H.J., Huang P.H.;  
RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
RT reovirus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1733;  
RA Liu H.J., Huang P.H.;  
RT "Molecular cloning and sequencing of the sigma A-encoded gene of Avian  
RT orthoreovirus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF294770; AAG44969.1; -.  
DR EMBL; AF293773; AAG44956.1; -.  
DR EMBL; AF294763; AAG44962.1; -.  
DR InterPro; IPR004317; Sigma\_1\_2.  
DR Pfam; PF03084; Sigma\_1\_2; 1.  
SQ SEQUENCE 416 AA; 46106 MW; CDE90302CCF2C0FF CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : | |  
Db 189 CARLTQSC 196

RESULT 6

Q9DPG7

ID Q9DPG7 PRELIMINARY; PRT; 416 AA.  
AC Q9DPG7;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=T6;  
 RA Liu H.J., Huang P.H.;  
 RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
 RT reovirus.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF294768; AAG44967.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46134 MW; DF7088105579C0FF CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

#### RESULT 7

##### Q9DPG8

ID Q9DPG8 PRELIMINARY; PRT; 416 AA.  
 AC Q9DPG8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=750505;  
 RA Liu H.J., Huang P.H.;  
 RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
 RT reovirus.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF294767; AAG44966.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46034 MW; B8BAAD27161121FB CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

RESULT 8

O72459

ID O72459 PRELIMINARY; PRT; 416 AA.  
 AC O72459;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Major inner capsid protein sigma 1.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=176;  
 RX MEDLINE=99348515; PubMed=10417266;  
 RA Duncan R.;  
 RT "Extensive sequence divergence and phylogenetic relationships between  
 RT the fusogenic and nonfusogenic orthoreoviruses: A species proposal."; Virology 260:316-328(1999).  
 RL  
 DR EMBL; AF059716; AAC18121.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46090 MW; BB497A899F1121FB CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

RESULT 9

Q9DPH0

ID Q9DPH0 PRELIMINARY; PRT; 416 AA.  
 AC Q9DPH0;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R2/TW;  
 RA Liu H.J., Huang P.H.;  
 RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
 RT reovirus."; Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RL  
 DR EMBL; AF294765; AAG44964.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46070 MW; CC5DA45A1A9F3B41 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
| | : | : | |  
Db 189 CARLTQSC 196

RESULT 10

Q9DPH1

ID Q9DPH1 PRELIMINARY; PRT; 416 AA.  
AC Q9DPH1;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Sigma A.  
OS Avian orthoreovirus.  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=38170;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=916;  
RA Liu H.J., Huang P.H.;  
RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
RT reovirus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF294764; AAG44963.1; -.  
DR InterPro; IPR004317; Sigma\_1\_2.  
DR Pfam; PF03084; Sigma\_1\_2; 1.  
SQ SEQUENCE 416 AA; 46129 MW; 03B14CDF92D39F5D CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTRITESC 8  
| | : | : | |  
Db 189 CARLTQSC 196

RESULT 11

O72460

ID O72460 PRELIMINARY; PRT; 416 AA.  
AC O72460;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Major inner capsid protein sigma 1.  
OS Avian orthoreovirus.  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=38170;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=138;  
RX MEDLINE=99348515; PubMed=10417266;

RA Duncan R.;  
 RT "Extensive sequence divergence and phylogenetic relationships between  
 RT the fusogenic and nonfusogenic orthoreoviruses: A species proposal.";   
 RL Virology 260:316-328(1999).  
 DR EMBL; AF059717; AAC18122.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46065 MW; 68D9CE85C099C1F2 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

# RESULT 12

Q9YL31

ID Q9YL31 PRELIMINARY; PRT; 416 AA.  
 AC Q9YL31;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Major core protein sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S1133;  
 RX MEDLINE=20080971; PubMed=10612658;  
 RA Yin H.S., Shien J.H., Lee L.H.;  
 RT "Synthesis in Escherichia coli of avian reovirus core protein sigmaA  
 RT and its dsRNA-binding activity.";   
 RL Virology 266:33-41(2000).  
 DR EMBL; AF104311; AAD17921.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46148 MW; F7870C0CEE44A960 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

# RESULT 13

Q9DPG9

ID Q9DPG9 PRELIMINARY; PRT; 416 AA.  
 AC Q9DPG9;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=918;  
 RA Liu H.J., Huang P.H.;  
 RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
 RT reovirus.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF294766; AAG44965.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46082 MW; DA9F827068955ADF CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

#### RESULT 14

Q9E6F8

ID Q9E6F8 PRELIMINARY; PRT; 416 AA.  
 AC Q9E6F8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2408;  
 RA Liu H.H.J., Huang P.H., Chen J.H., Lin M.Y.;  
 RT "Molecular cloning and sequencing of the sigma A-encoding gene of  
 RT avian reovirus.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF247724; AAG09473.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46062 MW; B8BAB5B69F1121FB CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

RESULT 15

Q9DL59

ID Q9DL59 PRELIMINARY; PRT; 416 AA.  
 AC Q9DL59;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Sigma A.  
 OS Avian orthoreovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=38170;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=601G;  
 RA Liu H.J., Huang P.H., Chen J.H.;  
 RT "Molecular cloning and sequencing of the sigma A-encoded gene of avian  
 reovirus.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF311322; AAG45147.1; -.  
 DR InterPro; IPR004317; Sigma\_1\_2.  
 DR Pfam; PF03084; Sigma\_1\_2; 1.  
 SQ SEQUENCE 416 AA; 46066 MW; 0DBC223DE5245355 CRC64;

Query Match 78.3%; Score 36; DB 12; Length 416;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTRITESC 8  
 | | : | : | |  
 Db 189 CARLTQSC 196

Search completed: November 13, 2003, 09:51:09  
 Job time : 22.0833 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 70.6562 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-16  
Perfect score: 130  
Sequence: 1 WRCVLREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	130	100.0	21	18	AAW13415		Brain homing pepti
2	130	100.0	21	21	AAB12002		Brain homing pepti
3	130	100.0	21	22	AAE11808		Phage peptide #16
4	130	100.0	21	23	AAU10719		Brain homing pepti
5	130	100.0	21	24	ABU59531		Brain receptor tar
6	54	41.5	99	22	AAU44925		Propionibacterium
7	51.5	39.6	275	22	ABG14312		Novel human diagno
8	51.5	39.6	732	22	ABB61396		Drosophila melanog
9	51.5	39.6	873	22	AAE02339		Drosophila melanog
10	50	38.5	523	21	AAB15972		E. coli proliferat
11	49.5	38.1	67	22	ABG01611		Novel human diagno
12	49	37.7	62	19	AAW44771		Fragment of scorpi
13	49	37.7	84	19	AAW44774		T. stigmurus scorp
14	49	37.7	215	22	AAB63255		Human breast cance
15	48	36.9	83	22	AAU54782		Propionibacterium
16	48	36.9	415	22	ABG30150		Novel human diagno
17	48	36.9	482	23	ABB06017		Monascus purpureus
18	47	36.2	9	18	AAW13435		Brain homing pepti
19	47	36.2	9	21	AAB12006		Brain homing pepti
20	47	36.2	9	22	AAE11812		Phage peptide #20
21	47	36.2	9	23	AAU10723		Brain homing pepti
22	47	36.2	73	22	AAU46781		Propionibacterium
23	47	36.2	759	24	ABP97378		Human kielin-like
24	47	36.2	973	8	AAP70769		Glycoprotein B of
25	47	36.2	1057	24	ABP97370		Human kielin-like
26	47	36.2	1192	24	ABP97376		Human kielin-like
27	47	36.2	1207	24	ABP97377		Human kielin-like
28	47	36.2	1251	24	ABP97375		Human kielin-like
29	47	36.2	1342	24	ABP97379		Human kielin-like
30	47	36.2	1477	24	ABP97371		Human kielin-like
31	47	36.2	1512	24	ABP97372		Human kielin-like
32	47	36.2	1535	24	ABP97374		Human kielin-like
33	47	36.2	1570	24	ABP97373		Human kielin-like
34	47	36.2	1593	24	ABP97369		Human kielin-like
35	47	36.2	1628	24	ABP97368		Human kielin-like
36	46.5	35.8	68	22	AAM80357		Human haematologic
37	46.5	35.8	91	22	AAG65621		Novel human protei
38	46.5	35.8	120	22	AAU41735		Propionibacterium
39	46.5	35.8	128	21	AAY86527		Human gene 72-enco
40	46.5	35.8	165	21	AAY86526		Human gene 72-enco
41	46	35.4	43	17	AAR95475		V4, monoclonal ant
42	46	35.4	67	22	AAE03499		Human gene 19 enco
43	46	35.4	67	23	ABG63344		Human albumin fusi
44	46	35.4	177	22	AAB65867		Human INTERCEPT 25
45	46	35.4	206	22	AAB65870		Human INTERCEPT 25

# ALIGNMENTS

RESULT 1

AAW13415

ID AAW13415 standard; Peptide; 21 AA.

XX

AC AAW13415;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN W09710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 13; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
CC peptide that was identified using a novel method for obtaining  
CC molecules that home to a selected organ or tissue. This in vivo  
CC panning method typically involves administering a phage display  
CC library to a subject, and identifying expressed peptides which  
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
CC vascular tissue or tumour tissue. The isolated peptides (see  
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
CC labels to the selected organ/tissue (claimed) or to identify and/or  
CC isolate target molecules (claimed). The peptides can be directly  
CC identified in vivo, as compared to prior art in vitro screening  
CC methods, which require further examination to see if they maintain  
CC specificity in vivo.

XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 130; DB 18; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.1e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHRL 21

Db                    |||||  
1 WRCVLREGPAGGCAWFNRHRL 21

RESULT 2

AAB12002

ID    AAB12002 standard; peptide; 21 AA.

XX

AC    AAB12002;

XX

DT    17-OCT-2000    (first entry)

XX

DE    Brain homing peptide # 16.

XX

KW    Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS    Mus sp.

XX

PN    US6068829-A.

XX

PD    30-MAY-2000.

XX

PF    23-JUN-1997;    97US-0862855.

XX

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Pasqualini R,   Ruoslahti E;

XX

DR    WPI; 2000-410850/35.

XX

PT    Identifying and recovering organ homing molecules or peptides by in

PT    vivo panning comprises administering a library of diverse peptides

PT    linked to a tag which facilitates recovery of these peptides -

XX

PS    Example 2; Column 17; 20pp; English.

XX

CC    The present sequence is a mouse brain homing peptide. This sequence was

CC    identified by using in vivo panning to screen a library of potential

CC    organ homing molecules. The present sequence can be used to direct a

CC    moiety to a the brain tissue, by linking the moiety to the present

CC    sequence. Examples of potential moieties are drugs, toxins or a

CC    detectable label. The present sequence contains a VRL amino acid motif.

XX

SQ    Sequence    21 AA;

Query Match                    100.0%;    Score 130;    DB 21;    Length 21;

Best Local Similarity    100.0%;    Pred. No. 1.1e-11;

Matches    21;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy                    1 WRCVLREGPAGGCAWFNRHRL 21

|||||

Db                    1 WRCVLREGPAGGCAWFNRHRL 21

RESULT 3

AAE11808

ID AAE11808 standard; peptide; 21 AA.

XX

AC AAE11808;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #16 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

FH Key Location/Qualifiers

FT Domain 4..6

FT /label= VLR\_motif

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo

PT panning that selectively home to a selected organ or tissue useful for

PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
 CC molecules that selectively home to a selected organ or tissue such as  
 CC brain, kidney or tumour recovered by in vivo panning. The invention  
 CC generally relates to the field of molecular medicine, drug delivery and  
 CC to a method of invivo panning for identifying a molecule that homes to a  
 CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
 CC and fragments of proteins contained in an enriched library fraction may  
 CC be administered to a subject as part of a pharmaceutical composition to  
 CC treat disease or in diagnostic methods. The present sequence is a  
 CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 130; DB 22; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.1e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRCVLREGPAGGCAWFNRHRL 21  
|||  
Db 1 WRCVLREGPAGGCAWFNRHRL 21

RESULT 4

AAU10719

ID AAU10719 standard; peptide; 21 AA.

XX

AC AAU10719;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #16 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety  
CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying

CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 130; DB 23; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHRL 21  
| | | | | | | | | | | | | | | | | | | | |  
Db 1 WRCVLREGPAGGCAWFNRHRL 21

# RESULT 5

ABU59531

ID ABU59531 standard; Peptide; 21 AA.

XX

AC ABU59531;

XX

DT 22-APR-2003 (first entry)

XX

DE Brain receptor targeting peptide #3.

XX

KW Targeting ligand; bioactive agent; polymer matrix; cancer; cytostatic;

KW cathepsin-D substrate; peptides; neuroreceptor; adrenal receptor;

KW fibronectin; vitronectin; integrin; RGD motif; angiogenic endothelium;

KW tumour; cationic cancer-targeting peptide.

XX

OS Synthetic.

XX

PN US2002041898-A1.

XX

PD 11-APR-2002.

XX

PF 25-JUL-2001; 2001US-0912609.

XX

PR 05-JAN-2000; 2000US-0478124.

PR 31-OCT-2000; 2000US-0703474.

XX

PA (UNGE/) UNGER E C.

PA (MATS/) MATSUNAGA T O.

PA (RAMA/) RAMASWAMI V.

PA (ROMA/) ROMANOWSKI M J.

XX

PI Unger EC, Matsunaga TO, Ramaswami V, Romanowski MJ;

XX

DR WPI; 2003-208921/20.

XX

PT Targeted delivery system comprising a bioactive agent homogeneously

PT dispersed in a targeted matrix is especially useful in cancer therapy

PT -

XX

PS Claim 23; Page 37; 46pp; English.

XX

CC The invention relates to a composition comprising a bioactive agent  
CC homogeneously dispersed in a targeted matrix (polymer and targeting  
CC ligand). Also included are a targeted matrix for use as a delivery  
CC vehicle comprising a polymer associated with a targeting ligand,  
CC enhancing the bioavailability of an agent comprising administration  
CC of the composition and treating cancer comprising administration of the  
CC novel composition. The method is useful for targeted delivery of a drug,  
CC especially in cancer therapy. The targeting ligand may be a peptide.  
CC Examples of targeting peptides are disclosed including cathepsin-D  
CC substrate peptides, peptides targeting receptors in the brain and  
CC kidney, peptides recognising fibronectin- and vitronectin-binding  
CC integrins, peptides targeting the RGD (Arg-Gly-Asp)-motif in, e.g.,  
CC antibodies, peptides targeting the angiogenic endothelium of solid  
CC tumours, tissue specific peptides (e.g. of lung, skin, pancreas,  
CC intestine, uterus, adrenal gland and retina), and cationic cancer-  
CC targeting peptides. The present sequence is a peptide targeting  
CC ligand disclosed in the invention.

XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 130; DB 24; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHRL 21

|||||||

Db 1 WRCVLREGPAGGCAWFNRHRL 21

#### RESULT 6

AAU44925

ID AAU44925 standard; Protein; 99 AA.

XX

AC AAU44925;

XX

DT 27-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #5821.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

XX

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US12865.

XX

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.



XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59524.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID No 6120; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 99 AA;

Query Match 41.5%; Score 54; DB 22; Length 99;  
Best Local Similarity 60.0%; Pred. No. 3.6;  
Matches 12; Conservative 0; Mismatches 2; Indels 6; Gaps 2;

Qy 1 WRCVLREGPAGGCAWFNRHR 20  
|| || || || || ||  
Db 6 WR--LRSGPTGGC----RHR 19

RESULT 7  
ABG14312  
ID ABG14312 standard; Protein; 275 AA.  
XX  
AC ABG14312;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #14303.  
XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS78499.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 44671; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 275 AA;

Query Match 39.6%; Score 51.5; DB 22; Length 275;  
 Best Local Similarity 57.9%; Pred. No. 23;  
 Matches 11; Conservative 1; Mismatches 2; Indels 5; Gaps 1;

Qy 5 LREGPAGGCA-----WFNR 18

Db                   | | | | | | | |       | : |  
111 LPEGPAGGCAQNPGLWASR 129

RESULT 8

ABB61396

ID   ABB61396 standard; Protein; 732 AA.

XX

AC   ABB61396;

XX

DT   26-MAR-2002   (first entry)

XX

DE   Drosophila melanogaster polypeptide SEQ ID NO 10980.

XX

KW   Drosophila; developmental biology; cell signalling; insecticide;  
KW   pharmaceutical.

XX

OS   Drosophila melanogaster.

XX

PN   WO200171042-A2.

XX

PD   27-SEP-2001.

XX

PF   23-MAR-2001; 2001WO-US09231.

XX

PR   23-MAR-2000; 2000US-191637P.

PR   11-JUL-2000; 2000US-0614150.

XX

PA   (PEKE ) PE CORP NY.

XX

PI   Venter JC,   Adams M,   Li PWD,   Myers EW;

XX

DR   WPI; 2001-656860/75.

DR   N-PSDB; ABL05499.

XX

PT   New isolated nucleic acid detection reagent for detecting 1000 or more  
PT   genes from Drosophila and for elucidating cell signalling and cell-cell  
PT   interactions -

XX

PS   Disclosure; SEQ ID NO 10980; 21pp + Sequence Listing; English.

XX

CC   The invention relates to an isolated nucleic acid detection reagent  
CC   capable of detecting 1000 or more genes from Drosophila. The invention is  
CC   useful in developmental biology and in elucidating cell signalling and  
CC   cell-cell interactions in higher eukaryotes for the development of  
CC   insecticides, therapeutics and pharmaceutical drugs. The invention  
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC   sequences (ABL01840-ABL16175) and the encoded proteins  
CC   (ABB57737-ABB72072).

CC   The sequence data for this patent did not form part of the printed  
CC   specification, but was obtained in electronic format directly from WIPO  
CC   at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ   Sequence    732 AA;

Query Match                   39.6%;   Score 51.5;   DB 22;   Length 732;  
Best Local Similarity       27.9%;   Pred. No. 61;

Matches 12; Conservative 1; Mismatches 7; Indels 23; Gaps 1;

```
QY      1 WRCVLREGPAGGCA-----WFNRHR 20
          ||| | |||
Db      81 WLCVLLVGIAAGCVAGMVDIGASWMSDLKHGICPPAFWFNREQ 123
```

RESULT 9

AAE02339

ID AAE02339 standard; Protein; 873 AA.

XX

AC AAE02339;

XX

DT 10-AUG-2001 (first entry)

XX

DE Drosophila melanogaster chloride channel (dmCLC) protein.

XX

KW Chloride channel; dmCLC; metazoan invertebrate; biopesticide;  
KW therapeutic.

XX

OS Drosophila melanogaster.

XX

FH Key Location/Qualifiers

FT Domain

113..133

FT /label= Transmembrane\_domain

FT Domain

185..205

FT /label= Transmembrane\_domain

FT Domain

265..285

FT /label= Transmembrane\_domain

FT Domain

318..338

FT /label= Transmembrane\_domain

FT Domain

338..345

FT /label= GKxGPxxH\_motif

FT /note= "Conserved signature sequence for  
FT anion-selective ion pores"

FT Domain

341..361

FT /label= Transmembrane\_domain

FT Domain

375..395

FT /label= Transmembrane\_domain

FT Domain

409..429

FT /label= Transmembrane\_domain

FT Domain

446..466

FT /label= Transmembrane\_domain

FT Domain

485..505

FT /label= Transmembrane\_domain

FT Domain

558..578

FT /label= Transmembrane\_domain

FT Domain

581..601

FT /label= Transmembrane\_domain

FT Domain

624..644

FT /label= Transmembrane\_domain

FT Domain

654..674

FT /label= Transmembrane\_domain

FT Domain

719..773

FT /label= CBS\_domain

FT Domain

778..798

FT /label= Transmembrane\_domain

FT Domain 808..860  
 FT /label= CBS\_domain  
 XX  
 PN WO200138359-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 29-NOV-2000; 2000WO-US32816.  
 XX  
 PR 29-NOV-1999; 99US-0167807.  
 PR 31-JAN-2000; 2000US-0179167.  
 PR 01-MAR-2000; 2000US-0186561.  
 PR 22-MAR-2000; 2000US-0190968.  
 PR 22-MAR-2000; 2000US-0191400.  
 XX  
 PA (GENO-) GENOPTERA LLC.  
 XX  
 PI Ebens AJ, Francis-Lang H, Keegan KP, Stout TJ, Kellerman KA;  
 PI Torpey J;  
 XX  
 DR WPI; 2001-355882/37.  
 DR N-PSDB; AAD05207.  
 XX  
 PT Invertebrate receptor nucleic acids isolated from Drosophila  
 PT melanogaster which can be used to genetically modify metazoan  
 PT invertebrate organisms resulting in expression or mis-expression of the  
 PT receptor protein -  
 XX  
 PS Claim 10; Page 70-72; 76pp; English.  
 XX  
 CC The patent discloses invertebrate receptor nucleic acids and  
 CC proteins isolated from Drosophila melanogaster. The sequences  
 CC of the present invention are used to genetically modify metazoan  
 CC invertebrate organisms such as insects and worms, resulting in the  
 CC expression or mis-expression of the receptor protein. The nucleic  
 CC acid molecules of the invention are used as hybridisation probes, in  
 CC expression vectors and to modify a host cell or animal and therefore  
 CC provide new means of providing biopesticides. The genetically modified  
 CC organisms are used in screening assays to identify compounds that are  
 CC potential pesticidal agents or therapeutics that interact with the  
 CC receptor proteins.  
 CC The present sequence is Drosophila melanogaster chloride channel  
 CC (dmCLC) protein.  
 XX  
 SQ Sequence 873 AA;  
  
 Query Match 39.6%; Score 51.5; DB 22; Length 873;  
 Best Local Similarity 27.9%; Pred. No. 72;  
 Matches 12; Conservative 1; Mismatches 7; Indels 23; Gaps 1;  
  
 QY 1 WRCVLREGPAGGCA-----WFNRHR 20  
 | | | | | :  
 Db 188 WLCVLLVGIAAGCVAGMVDIGASWMSDLKHGICPPAFWFNREQ 230

ID AAB15972 standard; Protein; 523 AA.  
 XX  
 AC AAB15972;  
 XX  
 DT 05-OCT-2000 (first entry)  
 XX  
 DE E. coli proliferation associated protein sequence SEQ ID NO:329.  
 XX  
 KW Escherichia coli; E. coli; proliferation; inhibition; screening;  
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200044906-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 27-JAN-2000; 2000WO-US02200.  
 XX  
 PR 27-JAN-1999; 99US-0117405.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2000-514822/46.  
 DR N-PSDB; AAA65977.  
 XX  
 PT Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 XX  
 PS Claim 11; Page 246-247; 316pp; English.  
 XX  
 CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
 CC sequences derived from Escherichia coli which inhibit E. coli  
 CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent  
 CC nucleotide and protein sequences associated with E. coli proliferation.  
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation- required gene in a microorganism, by contacting  
 CC a microorganism with a proliferation-required gene activity inhibitory  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria.  
 XX  
 SQ Sequence 523 AA;

Query Match 38.5%; Score 50; DB 21; Length 523;  
 Best Local Similarity 64.7%; Pred. No. 71;  
 Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
 || |||| ||| ||

RESULT 11

ABG01611

ID ABG01611 standard; Protein; 67 AA.

XX

AC ABG01611;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #1602.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS65798.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX

PS Claim 20; SEQ ID No 31970; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 67 AA;

Query Match 38.1%; Score 49.5; DB 22; Length 67;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 10; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

Qy 5 LREGPAGGCAWF-NRHRL 21

|| | || |: |||

Db 12 LRRWPGAGCWWGRRHRL 29

# RESULT 12

AAW44771

ID AAW44771 standard; peptide; 62 AA.

XX

AC AAW44771;

XX

DT 10-NOV-1998 (first entry)

XX

DE Fragment of scorpion T. stigmurus gamma toxin.

XX

KW Toxin; scorpion; chromatography; protease; serum; immunisation; horse;  
KW poisoning; human.

XX

OS Tityus stigmurus.

XX

PN BR9505982-A.

XX

PD 23-DEC-1997.

XX

PF 21-DEC-1995; 95BR-0005982.

XX

PR 21-DEC-1995; 95BR-0005982.

XX

PA (BUTA-) FUNDACAO BUTANTAN.

XX

PI Becerril-Lujan B, Calderon-Aranda ES, Corona-Villegas M;  
PI Coronas-Valderrama FI, Fletcher PL, Lucas SM, Martin BM;  
PI Possani LD, Raw I, Zamudio-Zuniga F;

XX

DR WPI; 1998-052767/06.

XX

PT Anti-scorpion serum production - by isolating genes and DNA from  
PT toxins in scorpion poison

XX

PS Disclosure; Fig 2; 20pp; Portuguese.

XX

CC This sequence represents a fragment of the gamma-st toxin from the  
CC scorpion Tityus stigmurus. The sequence is a composite of fragments  
CC generated by proteolytic digestion of the isolated toxin by the  
CC proteases Staphylococcus aureus V8, chymotrypsin, trypsin,



CC endopeptidases Asp-N or Lys-C (see AAW44774 for full length sequence).  
 CC The toxins were isolated from the scorpions and separated by  
 CC chromatographic methods. Their toxicity was determined by injection  
 CC of the chromatographic fragments into animals and observing for adverse  
 CC effects e.g. paralysis or mortality. The lethal toxins or fragments  
 CC were cleaved with proteases and their amino acid sequences determined.  
 CC Primers and probes were designed and used to isolate the gene encoding  
 CC the toxins. The toxins can be produced recombinantly and used to  
 CC generate sera for immunising horses and treating poisoning of humans  
 CC stung by scorpions.

XX

SQ Sequence 62 AA;

Query Match 37.7%; Score 49; DB 19; Length 62;  
 Best Local Similarity 53.8%; Pred. No. 12;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
 | |::| :| |||  
 Db 28 CTLKKGSSGYCAW 40

# RESULT 13

AAW44774

ID AAW44774 standard; Protein; 84 AA.

XX

AC AAW44774;

XX

DT 10-NOV-1998 (first entry)

XX

DE T. stigmurus scorpion gamma toxin.

XX

KW Toxin; scorpion; chromatography; protease; serum; immunisation; horse;  
 KW poisoning; human.

XX

OS Tityus stigmurus.

XX

FH Key Location/Qualifiers

FT Peptide 1..19

FT /note= "signal peptide"

FT Protein 20..84

FT /note= "mature protein"

XX

PN BR9505982-A.

XX

PD 23-DEC-1997.

XX

PF 21-DEC-1995; 95BR-0005982.

XX

PR 21-DEC-1995; 95BR-0005982.

XX

PA (BUTA-) FUNDACAO BUTANTAN.

XX

PI Becerril-Lujan B, Calderon-Aranda ES, Corona-Villegas M;  
 PI Coronas-Valderrama FI, Fletcher PL, Lucas SM, Martin BM;  
 PI Possani LD, Raw I, Zamudio-Zuniga F;

XX

DR WPI; 1998-052767/06.  
 DR N-PSDB; AAV05896.  
 XX  
 PT Anti-scorpion serum production - by isolating genes and DNA from  
 PT toxins in scorpion poison  
 XX  
 PS Disclosure; Fig 5; 20pp; Portuguese.  
 XX  
 CC This sequence represents the gamma toxin from the scorpion Tityus  
 CC stigmurus. The coding sequence was isolated using primers and probes  
 CC designed based on the amino acid sequence of proteolytic fragments of  
 CC the purified toxin. The toxins were isolated from the scorpions and  
 CC separated by chromatographic methods. Their toxicity was determined by  
 CC injection of the chromatographic fragments into animals and observing  
 CC for adverse effects e.g. paralysis or mortality. The lethal toxins or  
 CC fragments were cleaved with proteases and their amino acid sequences  
 CC determined. Digestion of the isolated toxin was performed by proteases:  
 CC Staphylococcus aureus V8, chymotrypsin, trypsin, endopeptidases Asp-N or  
 CC Lys-C. The toxins can be produced recombinantly and used to generate  
 CC sera for immunising horses and treating poisoning of humans stung by  
 CC scorpions.  
 XX  
 SQ Sequence 84 AA;

Query Match 37.7%; Score 49; DB 19; Length 84;  
 Best Local Similarity 53.8%; Pred. No. 16;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
 | |::| :| |||  
 Db 47 CTLKKGSSGYCAW 59

RESULT 14  
 AAB63255  
 ID AAB63255 standard; Protein; 215 AA.  
 XX  
 AC AAB63255;  
 XX  
 DT 26-MAR-2001 (first entry)  
 XX  
 DE Human breast cancer associated antigen protein sequence SEQ ID NO:617.  
 XX  
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;  
 KW cancer associated antigen; cytostatic; cancer vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073801-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 26-MAY-2000; 2000WO-US14749.  
 XX  
 PR 28-MAY-1999; 99US-0136526.  
 PR 10-SEP-1999; 99US-0153454.  
 XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Obata Y;

XX

DR WPI; 2001-025274/03.

XX

PT Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PT cancer -

XX

PS Example 1; Page 484-485; 799pp; English.

XX

CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

CC represent nucleotide sequences encoding human breast, gastric and

CC prostate cancer associated antigen precursors (CAAP) respectively.

CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

CC represent human breast, gastric and prostate CAAP protein sequence

CC respectively. CAAPs have cytostatic activity and can be used in the

CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterised by expression of an abnormal amount of a protein,

CC e.g. cancer.

XX

SQ Sequence 215 AA;

Query Match 37.7%; Score 49; DB 22; Length 215;

Best Local Similarity 55.6%; Pred. No. 40;

Matches 10; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 4 VLREGPAGGCAWFNRHRL 21

| | | | | | | |

Db 61 VPRRQTAGGAVWGRRHRL 78

RESULT 15

AAU54782

ID AAU54782 standard; Protein; 83 AA.

XX

AC AAU54782;

XX

DT 27-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #15678.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

XX

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US12865.

```

XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59566.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 15977; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 83 AA;

Query Match          36.9%; Score 48; DB 22; Length 83;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 11 GGCAWFNRHR 20
   | | | | |
Db 27 GTCCWFGRHR 36

```

Search completed: November 13, 2003, 09:45:29  
Job time : 71.6562 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 43.5312 Seconds  
(without alignments)  
88.069 Million cell updates/sec

Title: US-09-228-866-16  
Perfect score: 130  
Sequence: 1 WRCVIREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	54.5	41.9	553	10	US-09-349-385-8	Sequence 8, Appli	
2	54	41.5	275	12	US-10-029-386-32579	Sequence 32579, A	
3	50	38.5	523	9	US-09-912-020-329	Sequence 329, App	

4	49.5	38.1	41	12	US-10-269-806-175	Sequence 175, App
5	49.5	38.1	46	12	US-10-269-806-187	Sequence 187, App
6	49.5	38.1	46	12	US-10-269-806-193	Sequence 193, App
7	49	37.7	192	12	US-10-002-631C-48	Sequence 48, Appl
8	49	37.7	273	15	US-10-012-140-34	Sequence 34, Appl
9	47	36.2	9	12	US-10-306-878-12	Sequence 12, Appl
10	47	36.2	759	15	US-10-189-971-22	Sequence 22, Appl
11	47	36.2	1057	15	US-10-189-971-6	Sequence 6, Appli
12	47	36.2	1190	12	US-10-053-662A-2	Sequence 2, Appli
13	47	36.2	1192	15	US-10-189-971-18	Sequence 18, Appl
14	47	36.2	1207	15	US-10-189-971-20	Sequence 20, Appl
15	47	36.2	1251	15	US-10-189-971-16	Sequence 16, Appl
16	47	36.2	1342	15	US-10-189-971-24	Sequence 24, Appl
17	47	36.2	1477	15	US-10-189-971-8	Sequence 8, Appli
18	47	36.2	1512	15	US-10-189-971-10	Sequence 10, Appl
19	47	36.2	1535	15	US-10-189-971-14	Sequence 14, Appl
20	47	36.2	1570	15	US-10-189-971-12	Sequence 12, Appl
21	47	36.2	1593	15	US-10-189-971-4	Sequence 4, Appli
22	47	36.2	1628	15	US-10-189-971-2	Sequence 2, Appli
23	46.5	35.8	68	10	US-09-796-692-721	Sequence 721, App
24	46.5	35.8	68	15	US-10-040-862-721	Sequence 721, App
25	46.5	35.8	91	9	US-09-813-290-7	Sequence 7, Appli
26	46.5	35.8	128	15	US-10-012-542-465	Sequence 465, App
27	46.5	35.8	165	15	US-10-012-542-464	Sequence 464, App
28	46	35.4	370	11	US-09-796-753-76	Sequence 76, Appl
29	46	35.4	1192	12	US-10-053-662A-32	Sequence 32, Appl
30	45	34.6	160	15	US-10-102-806-558	Sequence 558, App
31	44.5	34.2	130	15	US-10-156-761-12342	Sequence 12342, A
32	44.5	34.2	348	15	US-10-161-572-56	Sequence 56, Appl
33	44.5	34.2	348	15	US-10-161-572-57	Sequence 57, Appl
34	44.5	34.2	352	10	US-09-978-295A-612	Sequence 612, App
35	44.5	34.2	352	10	US-09-978-697-612	Sequence 612, App
36	44.5	34.2	352	10	US-09-978-192A-612	Sequence 612, App
37	44.5	34.2	352	10	US-09-999-832A-612	Sequence 612, App
38	44.5	34.2	352	11	US-09-978-189-612	Sequence 612, App
39	44.5	34.2	352	11	US-09-978-608A-612	Sequence 612, App
40	44.5	34.2	352	11	US-09-978-585A-612	Sequence 612, App
41	44.5	34.2	352	11	US-09-978-191A-612	Sequence 612, App
42	44.5	34.2	352	11	US-09-978-403A-612	Sequence 612, App
43	44.5	34.2	352	11	US-09-978-564A-612	Sequence 612, App
44	44.5	34.2	352	11	US-09-999-833A-612	Sequence 612, App
45	44.5	34.2	352	11	US-09-981-915A-612	Sequence 612, App

#### ALIGNMENTS

#### RESULT 1

US-09-349-385-8

; Sequence 8, Application US/09349385

; Patent No. US20020152495A1

; GENERAL INFORMATION:

; APPLICANT: Ito, Toshiro

; APPLICANT: Fromm, Michael

; APPLICANT: Meyerowitz, Elliot

; TITLE OF INVENTION: PLANTS HAVING SEEDLESS FRUIT

; FILE REFERENCE: MBI-0002

```
; CURRENT APPLICATION NUMBER: US/09/349,385
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 60/115,967
; EARLIER FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Pinus radiata
; FEATURE:
; OTHER INFORMATION: translation of SEQ ID NO: 9
US-09-349-385-8
```

```
Query Match          41.9%; Score 54.5; DB 10; Length 553;
Best Local Similarity 23.0%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 3; Indels 41; Gaps 2;
```

```
Qy      1 WRCVL---REGPA-----GGCAWFNRH 19
      | |||  ::|||  |||||  ||
Db      13 WVCVLPLFTKDGPAYFLHSSDDVSAWRQWPLYIALLIVAVCAVLVSWLSPGGCAWAGR 72

Qy      20 R 20
      :
Db      73 K 73
```

## RESULT 2

```
US-10-029-386-32579
; Sequence 32579, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32579
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z97055.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUATION 4.00e-12
US-10-029-386-32579
```

Query Match 41.5%; Score 54; DB 12; Length 275;  
Best Local Similarity 46.7%; Pred. No. 8;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAW 15  
|| : | : | | : |  
Db 193 WRTLRCRQAPCGTCSW 207

RESULT 3

US-09-912-020-329  
; Sequence 329, Application US/09912020  
; Patent No. US20020045592A1  
; GENERAL INFORMATION:  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; TITLE OF INVENTION: ESCHERICHIA COLI  
; FILE REFERENCE: ELITRA.001DV1  
; CURRENT APPLICATION NUMBER: US/09/912,020  
; CURRENT FILING DATE: 2001-07-23  
; PRIOR APPLICATION NUMBER: 09/492,709  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/117,405  
; PRIOR FILING DATE: 1999-01-27  
; NUMBER OF SEQ ID NOS: 485  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 329  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: E. Coli  
US-09-912-020-329

Query Match 38.5%; Score 50; DB 9; Length 523;  
Best Local Similarity 64.7%; Pred. No. 52;  
Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
|| |||| ||| ||  
Db 20 LRHMPAGGVWWFNVDRH 36

RESULT 4

US-10-269-806-175  
; Sequence 175, Application US/10269806  
; Publication No. US20030176352A1  
; GENERAL INFORMATION:  
; APPLICANT: Min, Hosung  
; APPLICANT: Sitney, Karen  
; APPLICANT: Hartley, Cynthia



; TITLE OF INVENTION: Peptides and Related Compounds Having Thrombopoietic Activity

; FILE REFERENCE: A-750  
; CURRENT APPLICATION NUMBER: US/10/269,806  
; CURRENT FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 199  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 175  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized Peptide Sequence  
US-10-269-806-175

Query Match 38.1%; Score 49.5; DB 12; Length 41;  
Best Local Similarity 50.0%; Pred. No. 6;  
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 1 W-RCVLREGPAGGCAW 15  
| :|| :| |||:  
Db 11 WLQCVRAKGGGGGCSW 26

#### RESULT 5

US-10-269-806-187

; Sequence 187, Application US/10269806  
; Publication No. US20030176352A1  
; GENERAL INFORMATION:  
; APPLICANT: Min, Hosung  
; APPLICANT: Sitney, Karen  
; APPLICANT: Hartley, Cynthia  
; TITLE OF INVENTION: Peptides and Related Compounds Having Thrombopoietic Activity

; FILE REFERENCE: A-750  
; CURRENT APPLICATION NUMBER: US/10/269,806  
; CURRENT FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 199  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 187  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized Peptide Sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: At position 1, Fc at N-terminus  
US-10-269-806-187

Query Match 38.1%; Score 49.5; DB 12; Length 46;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 1 W-RCVLREGPAGGCAW 15  
| :|| :| |||:  
Db 11 WLQCVRAKGGGGGCSW 26

Db 16 WLQCVRAKGGGGGCSW 31

RESULT 6

US-10-269-806-193

; Sequence 193, Application US/10269806  
; Publication No. US20030176352A1  
; GENERAL INFORMATION:  
; APPLICANT: Min, Hosung  
; APPLICANT: Sitney, Karen  
; APPLICANT: Hartley, Cynthia  
; TITLE OF INVENTION: Peptides and Related Compounds Having Thrombopoietic Activity  
; FILE REFERENCE: A-750  
; CURRENT APPLICATION NUMBER: US/10/269,806  
; CURRENT FILING DATE: 2002-10-10  
; NUMBER OF SEQ ID NOS: 199  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 193  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized Peptide Sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (47)..(47)  
; OTHER INFORMATION: At position 47, Fc at C-terminus  
US-10-269-806-193

Query Match 38.1%; Score 49.5; DB 12; Length 46;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 1 W-RCVLREGPAGGCAW 15  
| :|| :| |||:|  
Db 11 WLQCVRAKGGGGGCSW 26

RESULT 7

US-10-002-631C-48

; Sequence 48, Application US/10002631C  
; Publication No. US20030157486A1  
; GENERAL INFORMATION:  
; APPLICANT: Graff, Jonathon M.  
; APPLICANT: Muenster, Matthew  
; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES  
; FILE REFERENCE: A34943 090495.0243  
; CURRENT APPLICATION NUMBER: US/10/002,631C  
; CURRENT FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: 60/300,309  
; PRIOR FILING DATE: 2001-06-21  
; NUMBER OF SEQ ID NOS: 324  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 192  
; TYPE: PRT

; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (2)...(192)  
; OTHER INFORMATION: Xaa = any amino acid  
US-10-002-631C-48

Query Match 37.7%; Score 49; DB 12; Length 192;  
Best Local Similarity 47.6%; Pred. No. 29;  
Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHRL 21  
|||:| | | | |||  
Db 45 WRCLLAEXHGGKWPLFXIHRL 65

RESULT 8

US-10-012-140-34

; Sequence 34, Application US/10012140  
; Publication No. US20030009017A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Glucksmann, Maria A.  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND  
USES

; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; PRIOR FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus amino acid sequence  
US-10-012-140-34

Query Match 37.7%; Score 49; DB 15; Length 273;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 9; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

Qy 4 VLREGPAG----GCAWFN 17  
:| ||||| || | :  
Db 139 ILAEGPAGYGNEGCCWLS 156

RESULT 9

US-10-306-878-12

; Sequence 12, Application US/10306878  
; Publication No. US20030175819A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, John C.  
; APPLICANT: Guo, Bin  
; TITLE OF INVENTION: Methods for Identifying Modulators of  
; TITLE OF INVENTION: Apoptosis  
; FILE REFERENCE: P-LJ 5535  
; CURRENT APPLICATION NUMBER: US/10/306,878  
; CURRENT FILING DATE: 2002-11-27  
; PRIOR APPLICATION NUMBER: US 60/334,149  
; PRIOR FILING DATE: 2001-11-28  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct  
US-10-306-878-12

Query Match 36.2%; Score 47; DB 12; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VLREGPAGG 12  
| | | | | | | |  
Db 1 VLREGPAGG 9

RESULT 10

US-10-189-971-22

; Sequence 22, Application US/10189971  
; Publication No. US20030028907A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030028907A1e1 Human Kielin-like Proteins and  
Polynucleotides Encoding the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: LEX-0360-USA  
; CURRENT APPLICATION NUMBER: US/10/189,971  
; CURRENT FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/302,949  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/315,634  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 759  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-189-971-22

Query Match 36.2%; Score 47; DB 15; Length 759;  
Best Local Similarity 77.8%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EGPAGGCAW 15  
|||||  
Db 43 EGPAGSCEW 51

RESULT 11

US-10-189-971-6

; Sequence 6, Application US/10189971  
; Publication No. US20030028907A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and  
Polynucleotides Encoding the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: LEX-0360-USA  
; CURRENT APPLICATION NUMBER: US/10/189,971  
; CURRENT FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/302,949  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/315,634  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1057  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-189-971-6

Query Match 36.2%; Score 47; DB 15; Length 1057;  
Best Local Similarity 77.8%; Pred. No. 2.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EGPAGGCAW 15  
|||||  
Db 341 EGPAGSCEW 349

RESULT 12

US-10-053-662A-2

; Sequence 2, Application US/10053662A  
; Publication No. US20030143545A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandra Charlesworth  
; APPLICANT: Falvia Spirito  
; APPLICANT: Guerrino Meneguzzi  
; APPLICANT: John Baird  
; APPLICANT: Keith Linder  
; TITLE OF INVENTION: ISOLATION OF THE LAMININ Y2 GENE IN

; TITLE OF INVENTION: HORSES AND ITS USE IN DIAGNOSING JUNCTIONAL  
EPIDERMOLYSIS

; TITLE OF INVENTION: BULLOSA  
; FILE REFERENCE: p84us4  
; CURRENT APPLICATION NUMBER: US/10/053,662A  
; CURRENT FILING DATE: 2002-01-24  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1190  
; TYPE: PRT  
; ORGANISM: Equine  
; FEATURE:  
; OTHER INFORMATION:  
US-10-053-662A-2

Query Match 36.2%; Score 47; DB 12; Length 1190;  
Best Local Similarity 40.0%; Pred. No. 2.9e+02;  
Matches 8; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHR 20  
|: ||| | |||  
Db 218 WKAVQRNGSPAKLQWSQRHR 237

#### RESULT 13

US-10-189-971-18

; Sequence 18, Application US/10189971  
; Publication No. US20030028907A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and  
Polynucleotides Encoding the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: LEX-0360-USA  
; CURRENT APPLICATION NUMBER: US/10/189,971  
; CURRENT FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/302,949  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/315,634  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 1192  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-189-971-18

Query Match 36.2%; Score 47; DB 15; Length 1192;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EGPAGGCAW 15  
||||| |

Db 476 EGPAGSCEW 484

RESULT 14

US-10-189-971-20

; Sequence 20, Application US/10189971  
; Publication No. US20030028907A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and  
Polynucleotides Encoding the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: LEX-0360-USA  
; CURRENT APPLICATION NUMBER: US/10/189,971  
; CURRENT FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/302,949  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/315,634  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 1207  
; TYPE: PRT  
; ORGANISM: homo sapiens

US-10-189-971-20

Query Match 36.2%; Score 47; DB 15; Length 1207;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EGPAGGCAW 15  
| | | | | | |  
Db 491 EGPAGSCEW 499

RESULT 15

US-10-189-971-16

; Sequence 16, Application US/10189971  
; Publication No. US20030028907A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and  
Polynucleotides Encoding the  
; TITLE OF INVENTION: Same  
; FILE REFERENCE: LEX-0360-USA  
; CURRENT APPLICATION NUMBER: US/10/189,971  
; CURRENT FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/302,949  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/315,634  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 1251  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-189-971-16

Query Match 36.2%; Score 47; DB 15; Length 1251;  
Best Local Similarity 77.8%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EGPAGGCAW 15  
||||| |  
Db 535 EGPAGSCEW 543

Search completed: November 13, 2003, 09:58:29  
Job time : 44.5312 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 21.875 Seconds  
(without alignments)  
92.322 Million cell updates/sec

Title: US-09-228-866-16  
Perfect score: 130  
Sequence: 1 WRCVLRREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	54.5	41.9	553	2	T08114	cytochrome P450 -
2	53.5	41.2	621	2	E82768	conserved hypothet
3	50	38.5	300	1	S57749	SURF1 protein - hu
4	50	38.5	523	2	S47758	hypothetical 59.4K
5	50	38.5	523	2	H91180	probable proteinas
6	50	38.5	523	2	C86027	probable proteinas
7	49	37.7	62	2	A61484	toxin VI - Brazili
8	49	37.7	84	2	S62867	toxin gamma precur
9	49	37.7	906	2	G83156	probable transcrip
10	48.5	37.3	302	2	T50737	bacteriochlorophyl
11	48	36.9	494	2	JX0300	ubiquinol-cytochro
12	48	36.9	666	2	T35864	hypothetical prote
13	47	36.2	477	2	A86474	unknown protein [i
14	47	36.2	489	2	F96504	protein F9C16.29 [
15	47	36.2	903	1	VGBEB1	glycoprotein B pre
16	47	36.2	2174	2	E95965	hypothetical glyci
17	46	35.4	661	2	D85361	hypothetical prote
18	46	35.4	1192	2	S69000	laminin gamma 2 ch
19	45.5	35.0	788	2	S53923	probable membrane
20	45	34.6	84	2	S62868	toxin gamma precur
21	45	34.6	438	2	E96545	hypothetical prote
22	44.5	34.2	475	2	E90318	medium-chain-fatty
23	44	33.8	65	1	NTSREB	neurotoxin XI - sc
24	44	33.8	84	2	S21158	neurotoxin TsVII p
25	44	33.8	114	2	A35153	histidine triad pr
26	44	33.8	519	2	A45982	calcium channel be
27	44	33.8	813	2	T21192	hypothetical prote
28	43.5	33.5	434	2	S73331	hypothetical prote
29	43.5	33.5	1694	2	S50065	sialoadhesin - mou
30	43	33.1	194	2	C72614	hypothetical prote
31	43	33.1	319	2	T11592	hypothetical prote
32	43	33.1	328	1	DWSMGG	dTDPglucose 4,6-de
33	43	33.1	369	2	T40052	probable transcrip
34	43	33.1	557	2	G86357	Similar to auxin-i
35	43	33.1	652	2	T19783	hypothetical prote
36	43	33.1	1193	2	A44018	laminin B2t chain
37	42.5	32.7	336	1	A69416	pyruvate formate-l
38	42.5	32.7	370	2	A96741	hypothetical prote
39	42.5	32.7	457	2	AG2699	magnesium transpor
40	42.5	32.7	457	2	G97481	mg2+ transport pro
41	42.5	32.7	625	1	A43030	protein-tyrosine k
42	42	32.3	118	2	G34792	Ig heavy chain pre
43	42	32.3	177	2	E71327	hypothetical prote
44	42	32.3	212	2	G81827	conserved hypothet
45	42	32.3	212	2	B81049	conserved hypothet

## ALIGNMENTS

RESULT 1  
T08114

cytochrome P450 - Monterey pine  
 C;Species: Pinus radiata (Monterey pine)  
 C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 04-Mar-2000  
 C;Accession: T08114  
 R;Bishop-Hurley, S.L.; Walter, C.; Gardner, R.C.  
 submitted to the EMBL Data Library, February 1998  
 A;Description: Isolation and expression of abundant mRNAs during somatic  
 embryogenesis of Pinus radiata.  
 A;Reference number: Z16362  
 A;Accession: T08114  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-553 <BIS>  
 A;Cross-references: EMBL:AF049067; NID:g2935524; PIDN:AAC05148.1; PID:g2935525  
 C;Genetics:  
 A;Gene: PRE74  
 C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
 C;Keywords: heme; iron; metalloprotein  
 F;343-513/Domain: cytochrome P450 homology <P45>  
 F;491/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 41.9%; Score 54.5; DB 2; Length 553;  
 Best Local Similarity 23.0%; Pred. No. 3.5;  
 Matches 14; Conservative 3; Mismatches 3; Indels 41; Gaps 2;

Qy 1 WRCVL---REGPA-----GGCAWFNRH 19  
 | ||| :||| |||| ||  
 Db 13 WVCVLPLFTKDGPAYFLHSSDDVSAWRQWPLYIALLIVAVCAVLVSWLSPGGCAWAGRH 72  
 Qy 20 R 20  
 :  
 Db 73 K 73

RESULT 2  
 E82768  
 conserved hypothetical protein XF0752 [imported] - Xylella fastidiosa (strain  
 9a5c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C;Accession: E82768  
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for  
 Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A;Reference number: A82515; MUID:20365717; PMID:10910347  
 A;Note: for a complete list of authors see reference number A59328 below  
 A;Accession: E82768  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-621 <SIM>  
 A;Cross-references: GB:AE003916; GB:AE003849; NID:g9105626; PIDN:AAF83562.1;  
 GSPDB:GN00128; XFSC:XF0752  
 A;Experimental source: strain 9a5c  
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.;  
 Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros,  
 M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.;

Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0752

Query Match 41.2%; Score 53.5; DB 2; Length 621;  
Best Local Similarity 45.8%; Pred. No. 5.4;  
Matches 11; Conservative 1; Mismatches 9; Indels 3; Gaps 1;

Qy 1 WRCVLREGPAGGCAWFN---RHRL 21  
| | | | | | | | | |  
Db 560 WHSSYRRAPADGVAWYNPGCRQRL 583

RESULT 3

S57749

SURF1 protein - human

C;Species: Homo sapiens (man)

C;Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C;Accession: S57749

R;Lennard, A.; Gaston, K.; Fried, M.

submitted to the EMBL Data Library, July 1994

A;Description: The Surf-1 and Surf-2 genes and their essential bidirectional promoter elements are conserved between mouse and human.

A;Reference number: S57747

A;Accession: S57749

A;Molecule type: mRNA

A;Residues: 1-300 <LEN>

A;Cross-references: EMBL:Z35093; NID:g895848; PIDN:CAA84476.1; PID:g895849

C;Comment: This protein is thought to be involved in cytochrome c oxidase biogenesis. Mutations are associated with Leigh's syndrome, a severe neurological disorder characterized by cytochrome c oxidase deficiency.

C;Genetics:

A;Gene: GDB:SURF1  
A;Cross-references: GDB:6071094; OMIM:185620  
A;Map position: 9q33-9q34  
C;Superfamily: human SURF1 protein

Query Match 38.5%; Score 50; DB 1; Length 300;  
Best Local Similarity 60.0%; Pred. No. 9;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WRCVLRREGPAGGCAW 15  
|| ||| | |||  
Db 25 WRSVLRVSPRPGVAW 39

#### RESULT 4

S47758

hypothetical 59.4K protein (dctA-dppF intergenic region) - Escherichia coli (strain K-12)

N;Alternate names: hypothetical protein o523

C;Species: Escherichia coli

C;Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 01-Mar-2002

C;Accession: S47758; C65152

R;Plunkett, G.

submitted to the EMBL Data Library, March 1994

A;Reference number: S47666

A;Accession: S47758

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-523 <PLU>

A;Cross-references: EMBL:U00039; NID:g466582; PIDN:AAB18514.1; PID:g466675

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: C65152

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-523 <BLAT>

A;Cross-references: GB:AE000431; GB:U00096; NID:g1789957; PIDN:AAC76561.1;

PID:g1789958; UWGP:b3536

A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: yhjS

C;Superfamily: Escherichia coli hypothetical 59.4K protein (dctA-dppF intergenic region)

Query Match 38.5%; Score 50; DB 2; Length 523;  
Best Local Similarity 64.7%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
|| ||| ||| ||  
Db 20 LRHMPAGGVWWFNVDRH 36

# RESULT 5

H91180

probable proteinase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: H91180

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: H91180

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-523 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA37839.1; PID:g13363890; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs4416

C;Superfamily: Escherichia coli hypothetical 59.4K protein (dctA-dppF intergenic region)

Query Match 38.5%; Score 50; DB 2; Length 523;  
Best Local Similarity 64.7%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
|| ||| ||| ||  
Db 20 LRHMPAGGVWVFNVDHRH 36

# RESULT 6

C86027

probable proteinase yhjS [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C;Accession: C86027

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: C86027

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-523 <STO>

A;Cross-references: GB:AE005174; NID:g12518259; PIDN:AAG58679.1; GSPDB:GN00145; UWGP:Z4952

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: yhjS

C;Superfamily: Escherichia coli hypothetical 59.4K protein (dctA-dppF intergenic region)

Query Match 38.5%; Score 50; DB 2; Length 523;  
Best Local Similarity 64.7%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
|| ||| ||| ||  
Db 20 LRHMPAGGVWVFNVDRH 36

#### RESULT 7

A61484

toxin VI - Brazilian scorpion

C;Species: Tityus serrulatus (Brazilian scorpion)

C;Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 02-Jun-1995

C;Accession: A61484

R;Marangoni, S.; Ghiso, J.; Sampaio, S.V.; Arantes, E.C.; Giglio, J.R.;

Oliveira, B.; Frangione, B.

J. Protein Chem. 9, 595-601, 1990

A;Title: The complete amino acid sequence of toxin TsTX-VI isolated from the venom of the scorpion Tityus serrulatus.

A;Reference number: A61484; MUID:91197385; PMID:2085384

A;Accession: A61484

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-62 <MAR>

C;Comment: This venom protein does not act as a neurotoxin in mice.

C;Superfamily: scorpion neurotoxin

C;Keywords: disulfide bond; monomer; venom

Query Match 37.7%; Score 49; DB 2; Length 62;  
Best Local Similarity 53.8%; Pred. No. 3.1;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
| |::| :| |||  
Db 28 CTLKKGSSGYCAW 40

#### RESULT 8

S62867

toxin gamma precursor - Tityus stigmurus

C;Species: Tityus stigmurus

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Aug-1999

C;Accession: S62867; S62865

R;Becerril, B.; Corona, M.; Coronas, F.I.V.; Zamudio, F.; Calderon-Aranda, E.S.;

Fletcher Jr., P.L.; Martin, B.M.; Possani, L.D.

Biochem. J. 313, 753-760, 1996

A;Title: Toxic peptides and genes encoding toxin gamma of the Brazilian scorpions Tityus bahiensis and Tityus stigmurus.

A;Reference number: S62861; MUID:96190713; PMID:8611151

A;Accession: S62867

A;Molecule type: DNA

A;Residues: 1-84 <BEC>  
A;Accession: S62865  
A;Molecule type: protein  
A;Residues: 20-81 <BEW>  
C;Superfamily: scorpion neurotoxin  
C;Keywords: amidated carboxyl end; neurotoxin; venom  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-82/Product: toxin gamma #status predicted <MAT>  
F;31-81,35-57,43-62,47-64/Disulfide bonds: #status predicted  
F;81/Modified site: amidated carboxyl end (Cys) (amide in mature form from following glycine) #status predicted

Query Match 37.7%; Score 49; DB 2; Length 84;  
Best Local Similarity 53.8%; Pred. No. 4;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
| |::| :| |||  
Db 47 CTLKKGSSGYCAW 59

#### RESULT 9

G83156

probable transcription regulator PA3921 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: G83156

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: G83156

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-906 <STO>

A;Cross-references: GB:AE004809; GB:AE004091; NID:g9950097; PIDN:AAG07308.1;

GSPDB:GN00131; PASP:PA3921

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3921

Query Match 37.7%; Score 49; DB 2; Length 906;  
Best Local Similarity 45.0%; Pred. No. 34;  
Matches 9; Conservative 2; Mismatches 3; Indels 6; Gaps 1;

Qy 8 GPAGG-----CAWFNRHRL 21  
||: | | ||: || |  
Db 356 GPSAGSLHLRACGWFSRHGL 375

# RESULT 10

T50737

bacteriochlorophyll a synthase (EC 6.1.-.-) bchG [imported] - Rhodobacter sphaeroides

C;Species: Rhodobacter sphaeroides

C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 02-Sep-2000

C;Accession: T50737

R;Choudhary, M.; Kaplan, S.

Nucleic Acids Res. 28, 862-867, 2000

A;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1.

A;Reference number: Z25222; MUID:20115911; PMID:10648776

A;Accession: T50737

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-302 <CHO>

A;Cross-references: EMBL:AF195122; PIDN:AAF24281.1

A;Experimental source: strain 2.4.1

C;Genetics:

A;Gene: bchG

C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0279

C;Keywords: ligase

Query Match 37.3%; Score 48.5; DB 2; Length 302;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 2 RCVLREGPAGGCAWFN 17  
| :||: ||| | |:|  
Db 263 RVLLRD-PAGKCPWYN 277

# RESULT 11

JX0300

ubiquinol-cytochrome-c reductase (EC 1.10.2.2) chain I precursor - Euglena gracilis mitochondrion

N;Alternate names: core 1 protein; mitochondrial enzyme; ubiquinol-cytochrome c oxidoreductase

C;Species: mitochondrion Euglena gracilis

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 03-Jun-2002

C;Accession: JX0300

R;Cui, J.Y.; Mukai, K.; Saeki, K.; Matsubara, H.

J. Biochem. 115, 98-107, 1994

A;Title: Molecular cloning and nucleotide sequences of cDNAs encoding subunits I, II, and IX of Euglena gracilis mitochondrial complex III.

A;Reference number: JX0301; MUID:94245672; PMID:8188644

A;Accession: JX0300

A;Molecule type: mRNA

A;Residues: 1-494 <CUI>

A;Cross-references: GB:D16671; NID:g464152; PIDN:BAA04079.1; PID:g464153

A;Note: this protein shows similarity to the members of the protein family which comprises complex III core proteins, mitochondrial processing peptidases and processing enhancing proteins

C;Comment: This protein plays an important role in electron transport and energy generation in mitochondrial inner membranes and some bacterial cell membranes.

C;Genetics:

A;Genome: mitochondrion



C;Superfamily: mitochondrial processing peptidase alpha chain  
C;Keywords: mitochondrion; oxidoreductase; respiratory chain  
F;1-18/Domain: propeptide #status predicted <PRO>  
F;18-494/Product: ubiquinol-cytochrome-c reductase chain I #status predicted  
<MAT>

Query Match 36.9%; Score 48; DB 2; Length 494;  
Best Local Similarity 55.6%; Pred. No. 27;  
Matches 10; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy 2 RCVLREGPAGGCAWFNRH 19  
| :||:| | | | |  
Db 443 RVLLRQGPRGGGDW--RH 458

RESULT 12

T35864

hypothetical protein SC9B1.19 - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C;Accession: T35864

R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.;  
Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A;Reference number: Z21591

A;Accession: T35864

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-666 <SAU>

A;Cross-references: EMBL:AL049727; PIDN:CAB41565.1; GSPDB:GN00070;

SCOEDB:SC9B1.19

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC9B1.19

Query Match 36.9%; Score 48; DB 2; Length 666;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNR 18  
| | : || |||| : |  
Db 69 WGRVPLDRPARGCAWADR 86

RESULT 13

A86474

unknown protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: A86474

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: A86474  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-477 <STO>  
A;Cross-references: GB:AE005172; NID:g11034948; PIDN:AAG27105.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 36.2%; Score 47; DB 2; Length 477;  
Best Local Similarity 45.0%; Pred. No. 37;  
Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 WRCVLREGPAGGCAWFNRHR 20  
| | | | | | | | | |  
Db 210 WTCVLSPIPRPKTEWFTRDR 229

#### RESULT 14

F96504

protein F9C16.29 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: F96504

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96504

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-489 <STO>  
A;Cross-references: GB:AE005173; NID:g8778668; PIDN:AAF79676.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F9C16.29  
A;Map position: 1

Query Match 36.2%; Score 47; DB 2; Length 489;  
Best Local Similarity 45.0%; Pred. No. 38;  
Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      1 WRCVLREGPAGGCAWFNRHR 20
          ||| | || |
Db     157 WTCVLSPIPRPKTEWFTRDR 176
```

#### RESULT 15

VGEBE1

glycoprotein B precursor - human herpesvirus 1 (strain F)

C;Species: human herpesvirus 1

C;Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C;Accession: A03750

R;Pellett, P.E.; Kousoulas, K.G.; Pereira, L.; Roizman, B.

J. Virol. 53, 243-253, 1985

A;Title: Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene:  
primary sequence and predicted protein structure of the wild type and of  
monoclonal antibody-resistant mutants.

A;Reference number: A03750; MUID:85083254; PMID:2981343

A;Accession: A03750

A;Molecule type: DNA

A;Residues: 1-903 <PEL>

A;Cross-references: GB:M14164; GB:M12398; NID:g330084; PIDN:AAA45776.1;

PID:g330086

C;Superfamily: herpesvirus glycoprotein B

C;Keywords: glycoprotein; transmembrane protein

F;1-29/Domain: signal sequence #status predicted <SIG>

F;30-903/Product: glycoprotein B #status predicted <MAT>

F;726-746/Domain: transmembrane #status predicted <TM1>

F;751-771/Domain: transmembrane #status predicted <TM2>

F;774-794/Domain: transmembrane #status predicted <TM3>

F;86,140,254,397,429,477,488,673,818,887/Binding site: carbohydrate (Asn)  
(covalent) #status predicted

F;115-572,132-528,206-270,363-411,595-632/Disulfide bonds: #status predicted

Query Match 36.2%; Score 47; DB 1; Length 903;  
Best Local Similarity 58.3%; Pred. No. 65;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy      5 LREGPAGGCAWF 16
          :|:| || ||
Db      1 MRQGAARGCRWF 12
```

Search completed: November 13, 2003, 09:53:02  
Job time : 22.875 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 12.0312 Seconds  
(without alignments)  
82.083 Million cell updates/sec

Title: US-09-228-866-16  
Perfect score: 130  
Sequence: 1 WRCVLREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54.5	41.9	553	1	CP78_PINRA	O65012 pinus radia
2	50.5	38.8	216	1	NOG1_BRARE	Q9w741 brachydanio
3	50	38.5	300	1	SUR1_HUMAN	Q15526 homo sapien
4	50	38.5	523	1	YHJS_ECOLI	P37657 escherichia
5	49	37.7	62	1	TTX6_TITSE	P45669 tityus serr
6	49	37.7	84	1	NTXP_TITSE	O77463 tityus serr
7	49	37.7	84	1	SCX7_TITST	P56612 tityus stig
8	48.5	37.3	302	1	BCHG_RHOSH	Q9z5d6 rhodobacter
9	48	36.9	494	1	UCR1_EUGGR	P43264 euglena gra
10	47.5	36.5	223	1	NOG3_BRARE	Q9y hv3 brachydanio
11	47	36.2	903	1	VGLB_HSV1F	P06436 herpes simp
12	46	35.4	488	1	RNF8_MOUSE	Q8vc56 mus musculu
13	46	35.4	1191	1	LMG2_MOUSE	Q61092 mus musculu
14	45.5	35.0	788	1	YG4C_YEAST	P42935 saccharomyc
15	45	34.6	84	1	SCX7_TITBA	P56611 tityus bahi
16	44.5	34.2	348	1	KILO_RAT	Q9z0j8 rattus norv
17	44.5	34.2	411	1	PCL_RHOSH	O54075 rhodobacter
18	44.5	34.2	882	1	CT1B_FUSSO	P52959 fusarium so
19	44.5	34.2	904	1	VGLB_HSV1P	P08665 herpes simp
20	44	33.8	65	1	SCXB_BUTOC	P01486 buthus occi

21	44	33.8	84	1	SCX7_TITSE	P15226	tityus serr
22	44	33.8	114	1	YHIT_SYNP7	P32084	synechococc
23	44	33.8	485	1	RNF8_HUMAN	O76064	homo sapien
24	44	33.8	813	1	YTQJ_CAEEL	Q19673	caenorhabdi
25	43.5	33.5	434	1	YB49_MYCPN	P75037	mycoplasma
26	43.5	33.5	692	1	ANMX_HUMAN	Q9nvm4	homo sapien
27	43.5	33.5	1694	1	SN_MOUSE	Q62230	mus musculu
28	43	33.1	319	1	YDFC_SCHPO	Q10484	schizosacch
29	43	33.1	328	1	STRE_STRGR	P29782	streptomyce
30	43	33.1	1193	1	LMG2_HUMAN	Q13753	homo sapien
31	42.5	32.7	625	1	ITK_MOUSE	Q03526	mus musculu
32	42	32.3	177	1	Y415_TREPA	O83430	treponema p
33	42	32.3	400	1	DDX1_DROVI	Q24731	drosophila
34	42	32.3	405	1	DCP2_PEA	P51851	pisum sativ
35	42	32.3	463	1	IFT1_MOUSE	Q64282	mus musculu
36	42	32.3	504	1	ATIN_HSVBP	P30020	bovine herp
37	42	32.3	641	1	SCAB_RABIT	O97742	oryctolagus
38	42	32.3	657	1	YH09_RALSO	Q8xyp9	ralstonia s
39	42	32.3	746	1	CLC5_HUMAN	P51795	homo sapien
40	42	32.3	746	1	CLC5_MOUSE	Q9wvd4	mus musculu
41	42	32.3	746	1	CLC5_RAT	P51796	rattus norv
42	42	32.3	1281	1	IP3S_MOUSE	Q9z329	mus musculu
43	42	32.3	2701	1	IP3S_HUMAN	Q14571	homo sapien
44	42	32.3	2701	1	IP3S_RAT	P29995	rattus norv
45	41.5	31.9	198	1	CD8A_PONPY	P30433	pongo pygma

#### ALIGNMENTS

#### RESULT 1

CP78\_PINRA

ID CP78\_PINRA STANDARD; PRT; 553 AA.

AC O65012;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cytochrome P450 78A4 (EC 1.14.-.-).

GN CYP78A4 OR PRE74.

OS Pinus radiata (Monterey pine).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.

OX NCBI\_TaxID=3347;

RN [1]

RP SEQUENCE FROM N.A.

RA Bishop-Hurley S.L., Walter C., Gardner R.C.;

RT "Isolation and expression of abundant mRNAs during somatic

RT embryogenesis of Pinus radiata.";

RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.

CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AF049067; AAC05148.1; -.  
 DR PIR; T08114; T08114.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Heme.  
 FT METAL 491 491 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 553 AA; 62026 MW; FC4ED38BAD264018 CRC64;

Query Match 41.9%; Score 54.5; DB 1; Length 553;  
 Best Local Similarity 23.0%; Pred. No. 1.4;  
 Matches 14; Conservative 3; Mismatches 3; Indels 41; Gaps 2;

Qy 1 WRCVL---REGPA-----GGCAWFNRH 19  
 | ||| :||| ||||| ||  
 Db 13 WVCVLPLFTKDGPAYFLHSSDDVSAWRQWPLYIALLLIVAVCAVLVSWLSPGGCAWAGRH 72  
 Qy 20 R 20  
 :  
 Db 73 K 73

## RESULT 2

### NOG1\_BRARE

ID NOG1\_BRARE STANDARD; PRT; 216 AA.  
 AC Q9W741;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Noggin 1 precursor.  
 GN NOG1.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99423658; PubMed=10491267;  
 RA Fuerthauer M., Thisse B., Thisse C.;  
 RT "Three different noggin genes antagonize the activity of bone  
 RT morphogenetic proteins in the zebrafish embryo."  
 RL Dev. Biol. 214:181-196(1999).  
 CC -!- FUNCTION: INHIBITOR OF BONE MORPHOGENETIC PROTEINS (BMP)  
 CC SIGNALING. MAY PLAY AN IMPORTANT ROLE IN THE DORSOVENTRAL  
 CC PATTERNING OF THE EMBRYO.  
 CC -!- SUBUNIT: Homodimer; disulfide-linked (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DEVELOPMENTAL STAGE: DETECTED FOLLOWING THE ACTIVATION OF THE  
 CC ZYGOTIC GENOME IN A FEW DEEP CELLS OF THE MARGINAL REGION OF THE  
 CC BLASTODERM. FROM THE 5-12 SOMITE STAGE, EXPRESSION IS OBSERVED IN  
 CC THE DORSAL TELEENCEPHALON AND IN POSTERIOR AND VENTRAL PARTS OF THE  
 CC EYE FIELD. BY THE 12-SOMITE STAGE DETECTED ALL ALONG THE DORSAL  
 CC NEURAL TUBE FROM THE LEVEL OF THE DIENCEPHALON TO THE CAUDAL

CC SPINAL CORD AND THIS EXPRESSION PERSISTS UNTIL 24 HR OF  
 CC DEVELOPMENT. AT THE 15-SOMITE STAGE EXPRESSION IS SEEN IN THE  
 CC MIDLINE AROUND THE TAIL BUD. BETWEEN 15 AND 20 HR DEVELOPMENT  
 CC DORSAL AS WELL AS VENTRAL EXPRESSION IS OBSERVED IN RECENTLY  
 CC FORMED SOMITES WHILE IN MORE MATURE SOMITES, DETECTED ONLY  
 CC VENTRALLY. BY 24 HR DEVELOPMENT EXPRESSION IS LIMITED TO THE  
 CC VENTRAL SCLEROTOMAL ASPECT OF THE CAUDAL SOMITES. LATER IN  
 CC DEVELOPMENT DETECTED IN VERY RESTRICTED PARTS OF THE CNS.  
 CC -!- SIMILARITY: BELONGS TO THE NOGGIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF159147; AAD43132.1; -.  
 DR ZFIN; ZDB-GENE-991206-8; nogl.  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 216 NOGGIN 1.  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 216 AA; 25093 MW; 3108242F298ABBBE CRC64;

Query Match 38.8%; Score 50.5; DB 1; Length 216;  
 Best Local Similarity 66.7%; Pred. No. 2.2;  
 Matches 10; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 WRCVLREGPAGGCAW 15  
 ||||| | | | |||  
 Db 186 WRCVARRG-ALKCAW 199

### RESULT 3

#### SUR1\_HUMAN

ID SUR1\_HUMAN STANDARD; PRT; 300 AA.  
 AC Q15526;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Surfeit locus protein 1.  
 GN SURF1 OR SURF-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95217332; PubMed=7702754;  
 RA Lennard A., Gaston K., Fried M.;  
 RT "The Surf-1 and Surf-2 genes and their essential bidirectional  
 RT promoter elements are conserved between mouse and human."  
 RL DNA Cell Biol. 13:1117-1126(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Colon, Kidney, and Stomach;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP POSSIBLE FUNCTION, AND INVOLVEMENT IN LS.  
RX MEDLINE=99057338; PubMed=9843204;  
RA Zhu Z., Yao J., Johns T., Fu K., de Bie I., Macmillan C.,  
RA Cuthbert A.P., Newbold R.F., Wang J., Chevrette M., Brown G.K.,  
RA Brown R.M., Shoubbridge E.A.;  
RT "SURF1, encoding a factor involved in the biogenesis of cytochrome c  
RT oxidase, is mutated in Leigh syndrome.";  
RL Nat. Genet. 20:337-343(1998).  
RN [4]  
RP REVIEW ON LS VARIANTS.  
RX MEDLINE=21217212; PubMed=11317352;  
RA Pequignot M.O., Dey R., Zeviani M., Tiranti V., Godinot C., Poyau A.,  
RA Sue C., Di Mauro S., Abitbol M., Marsac C.;  
RT "Mutations in the SURF1 gene associated with Leigh syndrome and  
RT cytochrome C oxidase deficiency.";  
RL Hum. Mutat. 17:374-381(2001).  
RN [5]  
RP VARIANTS LS GLU-124 AND THR-246, AND VARIANT HIS-202.  
RX MEDLINE=20208350; PubMed=10746561;  
RA Poyau A., Buchet K., Bouzidi M.F., Zabot M.-T., Echenne B., Yao J.,  
RA Shoubbridge E.A., Godinot C.;  
RT "Missense mutations in SURF1 associated with deficient cytochrome c  
RT oxidase assembly in Leigh syndrome patients.";  
RL Hum. Genet. 106:194-205(2000).  
RN [6]  
RP VARIANT LS ASP-274.  
RX MEDLINE=20112415; PubMed=10647889;  
RA Teraoka M., Yokoyama Y., Ninomiya S., Inoue C., Yamashita S.,  
RA Seino Y.;  
RT "Two novel mutations of SURF1 in Leigh syndrome with cytochrome c  
RT oxidase deficiency.";  
RL Hum. Genet. 105:560-563(1999).  
CC -!- FUNCTION: Probably involved in the biogenesis of the COX complex.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By



```

CC      similarity).
CC      -!- DISEASE: Defects in SURF1 are a cause of Leigh syndrome (LS)
CC      [MIM:256000]. LS is a severe neurological disorder characterized
CC      by bilaterally symmetrical necrotic lesions in subcortical brain
CC      regions that is commonly associated with systemic cytochrome c
CC      oxidase (COX) deficiency.
CC      -!- SIMILARITY: BELONGS TO THE SURF1 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z35093; CAA84476.1; -.
DR      EMBL; BC028314; AAH28314.1; -.
DR      PIR; S57749; S57749.
DR      Genew; HGNC:11474; SURF1.
DR      MIM; 185620; -.
DR      MIM; 256000; -.
DR      MIM; 220110; -.
DR      GO; GO:0005746; C:mitochondrial electron transport chain comp. . .; TAS.
DR      GO; GO:0004129; F:cytochrome c oxidase activity; TAS.
DR      GO; GO:0009060; P:aerobic respiration; TAS.
DR      GO; GO:0008535; P:cytochrome c oxidase biogenesis; TAS.
DR      GO; GO:0006118; P:electron transport; TAS.
DR      InterPro; IPR002994; Surf1.
DR      Pfam; PF02104; SURF1; 1.
DR      ProDom; PD024360; Surf1; 1.
DR      PROSITE; PS50895; SURF1; 1.
KW      Transmembrane; Mitochondrion; Inner membrane; Disease mutation;
KW      Polymorphism; Leigh syndrome.
FT      TRANSMEM      61      79      POTENTIAL.
FT      TRANSMEM      274     290     POTENTIAL.
FT      VARIANT       124     124     G -> E (in LS).
FT                                     /FTId=VAR_007450.
FT      VARIANT       124     124     G -> R (in LS).
FT                                     /FTId=VAR_015258.
FT      VARIANT       202     202     D -> H.
FT                                     /FTId=VAR_007451.
FT      VARIANT       246     246     I -> T (in LS).
FT                                     /FTId=VAR_007452.
FT      VARIANT       274     274     Y -> D (in LS).
FT                                     /FTId=VAR_015259.
SQ      SEQUENCE      300 AA;  33331 MW;  EC890EA48A0EDE7A CRC64;

```

```

Query Match          38.5%;  Score 50;  DB 1;  Length 300;
Best Local Similarity 60.0%;  Pred. No. 3.6;
Matches      9;  Conservative      0;  Mismatches      6;  Indels      0;  Gaps      0;

```

```

Qy      1 WRCVLREGPAGGCAW 15
        || ||| | | ||
Db      25 WRSVLRVSPRPGVAW 39

```

# RESULT 4

YHJS\_ECOLI

ID YHJS\_ECOLI STANDARD; PRT; 523 AA.  
 AC P37657;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yhjS.  
 GN YHJS OR B3536.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes."  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U00039; AAB18514.1; -.  
 DR EMBL; AE000431; AAC76561.1; -.  
 DR PIR; S47758; S47758.  
 DR EcoGene; EG12263; yhjS.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 523 AA; 59428 MW; 4241AF8CE7A9DC35 CRC64;

Query Match 38.5%; Score 50; DB 1; Length 523;  
 Best Local Similarity 64.7%; Pred. No. 6;  
 Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 5 LREGPAGGCAWFN--RH 19  
 || |||| ||| ||  
 Db 20 LRHMPAGGVWVFNVDRH 36

# RESULT 5

TTX6\_TITSE

ID TTX6\_TITSE STANDARD; PRT; 62 AA.  
 AC P45669;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Tityustoxin VI (TsTX-VI) (Toxin VI) (Ts VI).  
 OS Tityus serrulatus (Brazilian scorpion).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
 OC Buthoidea; Buthidae; Tityus.

OX NCBI\_TaxID=6887;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RX MEDLINE=91197385; PubMed=2085384;  
 RA Marangoni S., Ghiso J., Sampaio S.V., Arantes E.C., Giglio J.R.,  
 RA Oliveira B., Frangione B.;  
 RT "The complete amino acid sequence of toxin TsTX-VI isolated from the  
 RT venom of the scorpion *Tityus serrulatus*.";  
 RL J. Protein Chem. 9:595-601(1990).  
 CC -!- FUNCTION: Does not evoke the usual symptoms induced by the typical  
 CC neurotoxins of this venom, but only a generalized allergic  
 CC reaction.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.  
 CC ALPHA-TOXIN SUBFAMILY.  
 DR PIR; A61484; A61484.  
 DR HSSP; P01484; 1AHO.  
 DR InterPro; IPR003614; Knot1.  
 DR InterPro; IPR002061; Scorpion\_toxinL.  
 DR Pfam; PF00537; toxin\_3; 1.  
 DR ProDom; PD000908; Scorpion\_toxinL; 1.  
 DR SMART; SM00505; Knot1; 1.  
 KW Allergen; Amidation.  
 FT DISULFID 12 62 BY SIMILARITY.  
 FT DISULFID 16 38 BY SIMILARITY.  
 FT DISULFID 24 43 BY SIMILARITY.  
 FT DISULFID 28 45 BY SIMILARITY.  
 FT MOD\_RES 62 62 AMIDATION (PROBABLE).  
 SQ SEQUENCE 62 AA; 6717 MW; EFF355CDB1594839 CRC64;

Query Match 37.7%; Score 49; DB 1; Length 62;  
 Best Local Similarity 53.8%; Pred. No. 1.1;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
 | |::| :| |||  
 Db 28 CTLKKGSSGYCAW 40

# RESULT 6

## NTXP\_TITSE

ID NTXP\_TITSE STANDARD; PRT; 84 AA.  
 AC 077463;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Non-toxic protein NTxP precursor (TsNTxP).  
 GN NTXP.  
 OS *Tityus serrulatus* (Brazilian scorpion).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
 OC Buthoidea; Buthidae; *Tityus*.  
 OX NCBI\_TaxID=6887;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guatimosim S.C., Prado V.F., Diniz C.R., Chavez-Olortegui C.,

RA Kalapothakis E.;  
 RT "Molecular cloning and genomic analysis of TsnTxp: an immunogenic  
 RT protein from Tityus serrulatus scorpion venom.";   
 RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP FUNCTION.  
 RC TISSUE=Venom;  
 RX MEDLINE=97235459; PubMed=9080578;  
 RA Chavez-Olortegui C., Kalapothakis E., Ferreira A.M.B.M.,  
 RA Ferreira A.P., Diniz C.R.;  
 RT "Neutralizing capacity of antibodies elicited by a non-toxic protein  
 RT purified from the venom of the scorpion Tityus serrulatus.";   
 RL Toxicon 35:213-221(1997).  
 CC -!- FUNCTION: This protein is not toxic. It induces an immune response  
 CC similar to that induced by whole venom. Thus, polyclonal  
 CC antibodies raised against this protein can neutralize the effects  
 CC of the venom.  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.  
 CC ALPHA-TOXIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF039600; AAC25689.1; -.  
 DR EMBL; AF039599; AAC25688.1; -.  
 DR HSSP; P01484; 1AHO.  
 DR InterPro; IPR003614; Knot1.  
 DR InterPro; IPR002061; Scorpion\_toxinL.  
 DR Pfam; PF00537; toxin\_3; 1.  
 DR ProDom; PD000908; Scorpion\_toxinL; 1.  
 DR SMART; SM00505; Knot1; 1.  
 KW Signal; Amidation.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 FT CHAIN 20 81 NON-TOXIC PROTEIN NTXP.  
 FT MOD\_RES 81 81 AMIDATION (G-82 PROVIDE AMIDE GROUP)  
 FT (PROBABLE).  
 FT PROPEP 82 84  
 FT DISULFID 31 81 BY SIMILARITY.  
 FT DISULFID 35 57 BY SIMILARITY.  
 FT DISULFID 43 62 BY SIMILARITY.  
 FT DISULFID 47 64 BY SIMILARITY.  
 SQ SEQUENCE 84 AA; 9176 MW; DDEDE77B5B18C8EA CRC64;

Query Match 37.7%; Score 49; DB 1; Length 84;  
 Best Local Similarity 53.8%; Pred. No. 1.5;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
 | |::| :| |||  
 Db 47 CTLKKGSSGYCAW 59

RESULT 7

SCX7\_TITST

ID SCX7\_TITST STANDARD; PRT; 84 AA.  
 AC P56612;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Toxin gamma precursor.  
 OS Tityus stigmurus (Brazilian scorpion).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
 OC Buthoidea; Buthidae; Tityus.  
 OX NCBI\_TaxID=50344;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-81 FROM N.A.  
 RC TISSUE=Venom;  
 RX MEDLINE=96190713; PubMed=8611151;  
 RA Becerril B., Corona M., Coronas F.I., Zamudio F.,  
 RA Calderon-Aranda E.S., Fletcher P.L. Jr., Martin B.M., Possani L.D.;  
 RT "Toxic peptides and genes encoding toxin gamma of the Brazilian  
 RT scorpions Tityus bahiensis and Tityus stigmurus.";  
 RL Biochem. J. 313:753-760(1996).  
 CC -!- FUNCTION: Binds to sodium channels and inhibits the inactivation  
 CC of the activated channels, thereby blocking neuronal transmission.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.  
 CC BETA-TOXIN SUBFAMILY.  
 DR PIR; S62867; S62867.  
 DR HSSP; P01484; 1PTX.  
 DR InterPro; IPR003614; Knot1.  
 DR InterPro; IPR002061; Scorpion\_toxinL.  
 DR Pfam; PF00537; toxin\_3; 1.  
 DR ProDom; PD000908; Scorpion\_toxinL; 1.  
 DR SMART; SM00505; Knot1; 1.  
 KW Toxin; Neurotoxin; Ionic channel inhibitor; Sodium channel inhibitor;  
 KW Amidation; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 81 TOXIN GAMMA.  
 FT DISULFID 31 81 BY SIMILARITY.  
 FT DISULFID 35 57 BY SIMILARITY.  
 FT DISULFID 43 62 BY SIMILARITY.  
 FT DISULFID 47 64 BY SIMILARITY.  
 FT MOD\_RES 81 81 AMIDATION (G-82 PROVIDE AMIDE GROUP)  
 FT (PROBABLE).  
 SQ SEQUENCE 84 AA; 9366 MW; 460653ABAE1F7877 CRC64;

Query Match 37.7%; Score 49; DB 1; Length 84;  
 Best Local Similarity 53.8%; Pred. No. 1.5;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CVLREGPAGGCAW 15  
 | | : | : | | |  
 Db 47 CTLKKGSSGYCAW 59

# RESULT 8

BCHG\_RHOSH

ID BCHG\_RHOSH STANDARD; PRT; 302 AA.

AC Q9Z5D6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Bacteriochlorophyll synthase 33 kDa chain (Geranylgeranyl

DE bacteriochlorophyll synthase).

GN BCHG.

OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;

OC Rhodobacteraceae; Rhodobacter.

OX NCBI\_TaxID=1063;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;

RA Naylor G.W., Addlesee H.A., Gibson L.C.D., Hunter C.N.;

RT "The photosynthesis gene cluster of Rhodobacter sphaeroides.";

RL Photosyn. Res. 62:121-139(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;

RX MEDLINE=20115911; PubMed=10648776;

RA Choudhary M., Kaplan S.;

RT "DNA sequence analysis of the photosynthesis region of Rhodobacter

RT sphaeroides 2.4.1.";

RL Nucleic Acids Res. 28:862-867(2000).

CC -!- FUNCTION: CATALYZES THE ESTERIFICATION OF BACTERIOCHLOROPHYLLIDE A

CC BY GERANYLGERANIOL-PPI.

CC -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AJ010302; CAB38731.1; -.

DR EMBL; AF195122; AAF24281.1; -.

DR PIR; T50737; T50737.

DR InterPro; IPR006372; Chl\_synth.

DR InterPro; IPR000537; UbiA.

DR Pfam; PF01040; UbiA; 1.

DR TIGRFAMs; TIGR01476; chlor\_syn\_BchG; 1.

KW Photosynthesis; Bacteriochlorophyll biosynthesis; Transmembrane.

FT TRANSMEM 25 45 POTENTIAL.

FT TRANSMEM 49 69 POTENTIAL.

FT TRANSMEM 97 117 POTENTIAL.

FT TRANSMEM 119 139 POTENTIAL.

FT TRANSMEM 145 165 POTENTIAL.

FT TRANSMEM 166 186 POTENTIAL.

FT TRANSMEM 223 243 POTENTIAL.

FT TRANSMEM 246 266 POTENTIAL.

FT TRANSMEM 275 295 POTENTIAL.  
SQ SEQUENCE 302 AA; 32577 MW; A3EB487E3C900D42 CRC64;

Query Match 37.3%; Score 48.5; DB 1; Length 302;  
Best Local Similarity 56.2%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 2 RCVLREGPAGGCAWFN 17  
| :||: ||| | |:|  
Db 263 RVLLRD-PAGKCPWYN 277

#### RESULT 9

UCR1\_EUGGR

ID UCR1\_EUGGR STANDARD; PRT; 494 AA.

AC P43264;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial precursor (EC 1.10.2.2).

OS Euglena gracilis.

OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.

OX NCBI\_TaxID=3039;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SM-ZK;

RX MEDLINE=94245672; PubMed=8188644;

RA Cui J.-Y., Mukai K., Saeki K., Matsubara H.;

RT "Molecular cloning and nucleotide sequences of cDNAs encoding subunits I, II, and IX of Euglena gracilis mitochondrial complex III.";

RL J. Biochem. 115:98-107(1994).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY MEDiate FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.

CC -!- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2 ferrocytochrome c.

CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.

CC -!- PTM: THE N-TERMINUS IS BLOCKED.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D16671; BAA04079.1; -.

DR PIR; JX0300; JX0300.

DR MEROPS; M16.UNB; -.

DR InterPro; IPR001431; Peptidase\_M16.

DR Pfam; PF00675; Peptidase\_M16; 1.

DR Pfam; PF05193; Peptidase\_M16\_C; 1.

DR PROSITE; PS00143; INSULINASE; 1.  
 KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;  
 KW Oxidoreductase; Transit peptide; Zinc; Hydrolase; Metalloprotease.  
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).  
 FT CHAIN ? 494 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX  
 FT CORE PROTEIN I.  
 FT METAL 70 70 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 73 73 BY SIMILARITY.  
 FT METAL 74 74 ZINC (BY SIMILARITY).  
 FT METAL 150 150 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 494 AA; 54933 MW; 494D4C9AF74BDB9C CRC64;

Query Match 36.9%; Score 48; DB 1; Length 494;  
 Best Local Similarity 55.6%; Pred. No. 11;  
 Matches 10; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy 2 RCVLREGPAGGCAWFNRH 19  
 | : || : || | | | |  
 Db 443 RVLRLRQGPRGGGDW--RH 458

# RESULT 10

## NOG3\_BRARE

ID NOG3\_BRARE STANDARD; PRT; 223 AA.  
 AC Q9YHV3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Noggin 3 precursor.  
 GN NOG3.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=99102793; PubMed=9882485;  
 RA Bauer H., Meier A., Hild M., Stachel S., Economides A., Hazelett D.,  
 RA Harland R.M., Hammerschmidt M.;  
 RT "Follistatin and noggin are excluded from the zebrafish organizer.";  
 RL Dev. Biol. 204:488-507(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99423658; PubMed=10491267;  
 RA Fuerthauer M., Thisse B., Thisse C.;  
 RT "Three different noggin genes antagonize the activity of bone  
 RT morphogenetic proteins in the zebrafish embryo.";  
 RL Dev. Biol. 214:181-196(1999).  
 CC -!- FUNCTION: MAY FUNCTION AS AN INHIBITOR OF BONE MORPHOGENETIC  
 CC PROTEINS (BMP) SIGNALING DURING LATER STAGES OF DEVELOPMENT  
 CC INCLUDING LATE PHASES OF DORSOVENTRAL PATTERNING, TO REFINE THE  
 CC EARLY PATTERN SET UP BY THE INTERACTION OF CHORDINO AND BMP2/4.  
 CC NOT INVOLVED IN ORGANIZER FUNCTION OR EARLY PHASES OF DORSOVENTRAL  
 CC PATTERN FORMATION.  
 CC -!- SUBUNIT: Homodimer; disulfide-linked (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.



CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS LIMITED TO LATE STAGES OF  
 CC EMBRYOGENESIS. FIRST DETECTED AT 48 HRS OF DEVELOPMENT AND  
 CC RESTRICTED TO REGIONS OF ONGOING CHONDROGENESIS. EXPRESSION IS  
 CC OBSERVED IN THE ETHMOID PLATE AND THE TRABECULAE CRANII OF THE  
 CC NEUROCRANIUM AS WELL AS IN SOME PRESUMPTIVE CARTILAGE CELLS OF THE  
 CC PHARYNGEAL ARCHES. EXPRESSION IS FURTHERMORE OBSERVED IN THE  
 CC FORMING CARTILAGE OF THE PECTORAL FINS. AT 72 HRS OF DEVELOPMENT,  
 CC ACCUMULATES IN THE CERATOBANCHIAL AND BASIBRANCHIAL PARTS OF THE  
 CC GILL ARCHES.  
 CC -!- SIMILARITY: BELONGS TO THE NOGGIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF084949; AAD09176.1; -.  
 DR ZFIN; ZDB-GENE-990714-8; nog3.  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 223 NOGGIN 3.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 223 AA; 26029 MW; A21AE5DA36B75A37 CRC64;

Query Match 36.5%; Score 47.5; DB 1; Length 223;  
 Best Local Similarity 60.0%; Pred. No. 6.2;  
 Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 WRCVLREGPAGGCAW 15  
 |||| |:| |||  
 Db 193 WRCVQRKGGL-KCAW 206

RESULT 11  
 VGLB\_HSV1F  
 ID VGLB\_HSV1F STANDARD; PRT; 903 AA.  
 AC P06436;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB OR UL27.  
 OS Herpes simplex virus (type 1 / strain F).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85083254; PubMed=2981343;  
 RA Pellett P.E., Kousoulas K.G., Pereira L., Roizman B.;  
 RT "Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene:  
 RT primary sequence and predicted protein structure of the wild type and  
 RT of monoclonal antibody-resistant mutants.";

RL J. Virol. 53:243-253(1985).  
 RN [2]  
 RP SEQUENCE OF 1-176 FROM N.A.  
 RX MEDLINE=88306232; PubMed=2457278;  
 RA Hammerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,  
 RA Ludwig H.;  
 RT "Common epitopes of glycoprotein B map within the major DNA-binding  
 RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex  
 RT virus type 1 (HSV-1).";  
 RL Virology 165:406-418(1988).  
 CC -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
 CC GB, GC, GG, GD, GI, AND GE.  
 CC -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M14164; AAA45776.1; -.  
 DR EMBL; M21633; AAA45788.1; -.  
 DR PIR; A03750; VGBEB1.  
 DR InterPro; IPR000234; Glycoprot\_B.  
 DR Pfam; PF00606; Glycoprotein\_B; 1.  
 DR ProDom; PD000693; Glycoprot\_B; 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 903 GLYCOPROTEIN B.  
 FT DOMAIN 31 729 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 730 745 POTENTIAL.  
 FT TRANSMEM 751 770 POTENTIAL.  
 FT TRANSMEM 774 794 POTENTIAL.  
 FT DOMAIN 795 903 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 673 673 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 903 AA; 100104 MW; 73BDCA7813DB35E8 CRC64;

Query Match 36.2%; Score 47; DB 1; Length 903;  
 Best Local Similarity 58.3%; Pred. No. 28;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 5 LREGPAGGCAWF 16  
 :|:| | || ||  
 Db 1 MRQGAARGCRWF 12

RESULT 12

RNF8\_MOUSE  
 ID RNF8\_MOUSE STANDARD; PRT; 488 AA.  
 AC Q8VC56;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE RING finger protein 8.  
 GN RNF8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SIMILARITY: Contains 1 FHA domain.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; BC021778; AAH21778.1; -.  
 DR MGD; MGI:1929069; Rnf8.  
 DR InterPro; IPR000253; FHA.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS50006; FHA\_DOMAIN; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.

DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Zinc-finger.  
 FT DOMAIN 38 92 FHA.  
 FT DOMAIN 279 345 GLN-RICH.  
 FT ZN\_FING 406 444 RING-TYPE.  
 SQ SEQUENCE 488 AA; 55516 MW; 428242204EBC44A1 CRC64;

Query Match 35.4%; Score 46; DB 1; Length 488;  
 Best Local Similarity 34.5%; Pred. No. 22;  
 Matches 10; Conservative 2; Mismatches 7; Indels 10; Gaps 1;

QY 3 CVLREGPAG-----GCAWFNRHRL 21  
 |||:: || | || ||  
 Db 64 CVLKQNPEGQWTIMDNKSLNGVWLNRRERL 92

# RESULT 13

## LMG2\_MOUSE

ID LMG2\_MOUSE STANDARD; PRT; 1191 AA.  
 AC Q61092;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Laminin gamma-2 chain precursor (Kalinin/nicein/epiligrin 100 kDa subunit) (Laminin B2t chain).  
 GN LAMC2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB; TISSUE=Lung;  
 RX MEDLINE=95188894; PubMed=7882992;  
 RA Sugiyama S., Utani A., Yamada S., Kozak C.A., Yamada Y.;  
 RT "Cloning and expression of the mouse laminin gamma 2 (B2t) chain, a subunit of epithelial cell laminin."  
 RL Eur. J. Biochem. 228:120-128(1995).  
 RN [2]  
 RP REVISIONS.  
 RA Sasaki T., Yamada Y.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP BINDING TO HEPARIN; FIBULIN AND NIDOGEN, AND MUTAGENESIS OF ARG-76;  
 RP ARG-78; PHE-202; LYS-206; CYS-442 AND CYS-445.  
 RC STRAIN=FVB; TISSUE=Lung;  
 RX MEDLINE=21592560; PubMed=11733994;  
 RA Sasaki T., Goehring W., Mann K., Brakebusch C., Yamada Y.,  
 RA Faessler R., Timpl R.;  
 RT "Short arm region of laminin-5 gamma2 chain: structure, mechanism of processing and binding to heparin and proteins."  
 RL J. Mol. Biol. 314:751-763(2001).  
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.  
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three

different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end. The gamma-2 chain is a subunit of laminin-5 (epiligrin/kalinin/nicein) and binds fibulin-1, fibulin-1c, fibulin-2 and nidogen.

-!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

-!- TISSUE SPECIFICITY: EPITHELIAL CELLS OF MANY TISSUES, PARTICULARLY HIGH LEVELS IN TONGUE, HAIR FOLLICLES AND KIDNEY. BASEMENT MEMBRANES OF THE COLLECTING TUBULES OF KIDNEY AND PANCREAS.

-!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-!- DOMAIN: DOMAIN IV IS GLOBULAR.

-!- MISCELLANEOUS: Binds heparin.

-!- SIMILARITY: Contains 8 laminin EGF-like domains.

-!- SIMILARITY: Contains 1 laminin IV domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; U43327; AAA85256.2; -.  
DR HSSP; P02468; 1TLE.  
DR MGD; MGI:99913; Lamc2.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00052; laminin\_B; 1.  
DR Pfam; PF00053; laminin\_EGF; 5.  
DR PRINTS; PR00011; EGFLAMININ.  
DR SMART; SM00181; EGF; 7.  
DR SMART; SM00180; EGF\_Lam; 7.  
DR PROSITE; PS00022; EGF\_1; 4.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 6.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal;  
KW Heparin-binding.

FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	1191	LAMININ GAMMA-2 CHAIN.
FT	DOMAIN	28	83	LAMININ EGF-LIKE 1.
FT	DOMAIN	84	130	LAMININ EGF-LIKE 2.
FT	DOMAIN	139	186	LAMININ EGF-LIKE 3.
FT	DOMAIN	187	196	LAMININ EGF-LIKE 4 (N-TERMINAL).
FT	DOMAIN	197	381	LAMININ DOMAIN IV.
FT	DOMAIN	382	415	LAMININ EGF-LIKE 4 (C-TERMINAL).
FT	DOMAIN	416	461	LAMININ EGF-LIKE 5.
FT	DOMAIN	462	516	LAMININ EGF-LIKE 6.
FT	DOMAIN	517	572	LAMININ EGF-LIKE 7.
FT	DOMAIN	573	602	LAMININ EGF-LIKE 8 (INCOMPLETE).
FT	DOMAIN	603	1191	DOMAIN II AND I.
FT	DOMAIN	612	710	COILED COIL (POTENTIAL).

FT	DOMAIN	759	786	COILED COIL (POTENTIAL).
FT	DOMAIN	946	996	COILED COIL (POTENTIAL).
FT	DOMAIN	1139	1178	COILED COIL (POTENTIAL).
FT	SITE	586	588	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	84	96	BY SIMILARITY.
FT	DISULFID	86	102	BY SIMILARITY.
FT	DISULFID	104	113	BY SIMILARITY.
FT	DISULFID	116	128	BY SIMILARITY.
FT	DISULFID	139	150	BY SIMILARITY.
FT	DISULFID	141	155	BY SIMILARITY.
FT	DISULFID	157	166	BY SIMILARITY.
FT	DISULFID	169	184	BY SIMILARITY.
FT	DISULFID	462	470	BY SIMILARITY.
FT	DISULFID	464	481	BY SIMILARITY.
FT	DISULFID	484	493	BY SIMILARITY.
FT	DISULFID	496	514	BY SIMILARITY.
FT	DISULFID	517	531	BY SIMILARITY.
FT	DISULFID	519	538	BY SIMILARITY.
FT	DISULFID	541	550	BY SIMILARITY.
FT	DISULFID	553	570	BY SIMILARITY.
FT	DISULFID	610	610	INTERCHAIN (PROBABLE).
FT	DISULFID	613	613	INTERCHAIN (PROBABLE).
FT	DISULFID	1182	1182	INTERCHAIN (WITH BETA-3 CHAIN)
FT				(PROBABLE).
FT	CARBOHYD	342	342	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	362	362	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	526	526	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	941	941	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1032	1032	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MUTAGEN	76	76	R->A: NO CHANGE TO HERARIN-BINDING.
FT	MUTAGEN	78	78	R->A: NO CHANGE TO HERARIN-BINDING.
FT	MUTAGEN	202	202	F->A: NO FIBULIN-1C BINDING. NO CHANGE TO
FT				FIBULIN-2 BINDING.
FT	MUTAGEN	206	206	K->A: NO FIBULIN-1C BINDING. NO CHANGE TO
FT				FIBULIN-2 BINDING.
FT	MUTAGEN	442	442	C->S: 20-FOLD REDUCTION TO FIBULIN-2
FT				BINDING.
FT	MUTAGEN	445	445	C->S: 20-FOLD REDUCTION TO FIBULIN-2
FT				BINDING.
SQ	SEQUENCE	1191 AA;	130160 MW;	7016C1F851D909B9 CRC64;

Query Match 35.4%; Score 46; DB 1; Length 1191;  
 Best Local Similarity 40.0%; Pred. No. 50;  
 Matches 8; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHR 20  
 | : | | | | | | |  
 Db 218 WKAVQRNGAPAKLHWSQRHR 237

RESULT 14  
 YG4C\_YEAST  
 ID YG4C\_YEAST STANDARD; PRT; 788 AA.  
 AC P42935;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical 89.4 kDa Trp-Asp repeats containing protein in PMT6-PCT1  
DE intergenic region.  
GN YGR200C OR G7725.  
OS *Saccharomyces cerevisiae* (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=97060019; PubMed=8904340;  
RA Guerreiro P., Barreiros T., Soares H., Cyrne L., Maia e Silva A.,  
RA Rodrigues-Pousada C.;  
RT "Sequencing of a 17.6 kb segment on the right arm of yeast chromosome  
RT VII reveals 12 ORFs, including CCT, ADE3 and TR-I genes, homologues of  
RT the yeast PMT and EF1G genes, of the human and bacterial electron-  
RT transferring flavoproteins (beta-chain) and of the *Escherichia coli*  
RT phosphoserine phosphohydrolase, and five new ORFs.";  
RL Yeast 12:273-280(1996).  
CC -!- SIMILARITY: Contains 9 WD repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z49133; CAA88993.1; -.  
DR EMBL; Z72985; CAA97227.1; -.  
DR PIR; S53923; S53923.  
DR SGD; S0003432; YGR200C.  
DR GO; GO:0008023; C:transcription elongation factor complex; IDA.  
DR GO; GO:0016944; F:Pol II transcription elongation factor acti. . .; IMP.  
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; IMP.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 10.  
DR PRINTS; PR00320; GPROTEINBRPT.  
DR ProDom; PD000018; WD40; 1.  
DR SMART; SM00320; WD40; 12.  
DR PROSITE; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
DR PROSITE; PS50082; WD\_REPEATS\_2; 6.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 3.  
KW Hypothetical protein; Repeat; WD repeat.  
FT REPEAT 57 87 WD 1.  
FT REPEAT 101 130 WD 2.  
FT REPEAT 141 181 WD 3.  
FT REPEAT 200 234 WD 4.  
FT REPEAT 279 309 WD 5.  
FT REPEAT 383 413 WD 6.  
FT REPEAT 438 466 WD 7.  
FT REPEAT 604 634 WD 8.  
FT REPEAT 651 683 WD 9.  
SQ SEQUENCE 788 AA; 89410 MW; 5371908FE2E6EC0D CRC64;

Query Match

35.0%; Score 45.5; DB 1; Length 788;

Best Local Similarity 30.6%; Pred. No. 40;  
Matches 11; Conservative 2; Mismatches 8; Indels 15; Gaps 2;

Qy 1 WRCVLR-----EGPAGG---CAWFNRHRL 21  
| | | | : | | | | :  
Db 317 WVCSLR LGEMSSKGASTATGSSGGFWSC LWF THERM 352

# RESULT 15

## SCX7\_TITBA

ID SCX7\_TITBA STANDARD; PRT; 84 AA.  
AC P56611;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Toxin gamma precursor.  
OS Tityus bahiensis (Brazilian scorpion).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
OC Buthoidea; Buthidae; Tityus.  
OX NCBI\_TaxID=50343;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-81.  
RC TISSUE=Venom;  
RX MEDLINE=96190713; PubMed=8611151;  
RA Becerril B., Corona M., Coronas F.I., Zamudio F.,  
RA Calderon-Aranda E.S., Fletcher P.L. Jr., Martin B.M., Possani L.D.;  
RT "Toxic peptides and genes encoding toxin gamma of the Brazilian  
RT scorpions Tityus bahiensis and Tityus stigmurus.";  
RL Biochem. J. 313:753-760(1996).  
CC -!- FUNCTION: Binds to sodium channels and inhibits the inactivation  
CC of the activated channels, thereby blocking neuronal transmission.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
CC -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.  
CC BETA-TOXIN SUBFAMILY.  
DR PIR; S62868; S62868.  
DR HSSP; P01484; 1PTX.  
DR InterPro; IPR003614; Knot1.  
DR InterPro; IPR002061; Scorpion\_toxinL.  
DR Pfam; PF00537; toxin\_3; 1.  
DR ProDom; PD000908; Scorpion\_toxinL; 1.  
DR SMART; SM00505; Knot1; 1.  
KW Toxin; Neurotoxin; Ionic channel inhibitor; Sodium channel inhibitor;  
KW Amidation; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 81 TOXIN GAMMA.  
FT DISULFID 31 81 BY SIMILARITY.  
FT DISULFID 35 57 BY SIMILARITY.  
FT DISULFID 43 62 BY SIMILARITY.  
FT DISULFID 47 64 BY SIMILARITY.  
FT MOD\_RES 81 81 AMIDATION (G-82 PROVIDE AMIDE GROUP)  
FT (PROBABLE).  
SQ SEQUENCE 84 AA; 9384 MW; A24A2ACA7F768136 CRC64;

Query Match 34.6%; Score 45; DB 1; Length 84;  
Best Local Similarity 46.2%; Pred. No. 5.7;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;



Qy 3 CVLREGPAGGCAW 15  
| : : : | : | |||  
Db 47 CKIKKGSSGYCAW 59

Search completed: November 13, 2003, 09:46:39  
Job time : 13.0312 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 55.3438 Seconds  
(without alignments)  
97.917 Million cell updates/sec

Title: US-09-228-866-16  
Perfect score: 130  
Sequence: 1 WRCVLREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	58.5	45.0	627	16	Q8P8D7	Q8p8d7	xanthomonas
2	53.5	41.2	621	16	Q9PFC6	Q9pfc6	xylella fas
3	53.5	41.2	660	16	Q8PJW5	Q8pjw5	xanthomonas
4	51.5	39.6	788	5	Q9VUY1	Q9vuy1	drosophila
5	51.5	39.6	822	5	Q8IQN2	Q8iqn2	drosophila
6	51.5	39.6	882	5	Q8MQK2	Q8mqk2	drosophila
7	50	38.5	520	2	Q8GPQ9	Q8gpq9	pseudomonas
8	50	38.5	523	16	Q8X5L3	Q8x5l3	escherichia
9	50	38.5	523	16	Q8FCH0	Q8fch0	escherichia
10	50	38.5	1239	16	Q9FBZ4	Q9fbz4	streptomyce
11	49	37.7	245	16	Q8XT34	Q8xt34	ralstonia s
12	49	37.7	559	16	Q92ND1	Q92nd1	rhizobium m
13	49	37.7	906	16	Q9HX92	Q9hx92	pseudomonas
14	48	36.9	482	3	Q8X1W7	Q8x1w7	monascus an
15	48	36.9	666	16	Q9X838	Q9x838	streptomyce
16	48	36.9	1862	3	Q8J111	Q8j111	cryptococcu
17	47	36.2	237	10	Q8S038	Q8s038	oryza sativ
18	47	36.2	251	10	Q9C6X7	Q9c6x7	arabidopsis
19	47	36.2	358	12	Q8V5E1	Q8v5e1	ndelle viru
20	47	36.2	477	10	Q9C7I4	Q9c7i4	arabidopsis
21	47	36.2	489	10	Q9LNY9	Q9lny9	arabidopsis
22	47	36.2	903	12	Q69076	Q69076	human herpe
23	47	36.2	1160	12	Q9WP29	Q9wp29	bovine vira
24	47	36.2	1190	6	Q8HZI9	Q8hzi9	equus cabal
25	47	36.2	2174	16	Q92UU8	Q92uu8	rhizobium m
26	46.5	35.8	252	13	Q90568	Q90568	ginglymosto
27	46	35.4	126	13	Q9I839	Q9i839	anas platyr
28	46	35.4	126	13	Q9I882	Q9i882	gallus gall
29	46	35.4	179	2	Q8VS40	Q8vs40	klebsiella
30	46	35.4	275	11	Q9JK13	Q9jk13	mus musculu
31	46	35.4	464	11	Q61965	Q61965	mus musculu
32	46	35.4	661	10	Q65562	Q65562	arabidopsis
33	45.5	35.0	598	11	Q8R520	Q8r520	rattus norv
34	45	34.6	191	4	Q9BWP3	Q9bwp3	homo sapien
35	45	34.6	370	10	Q8GVN6	Q8gvn6	oryza sativ
36	45	34.6	438	10	Q9C6J4	Q9c6j4	arabidopsis
37	45	34.6	443	10	Q9FLS2	Q9fls2	arabidopsis
38	45	34.6	470	10	Q9AY54	Q9ay54	oryza sativ
39	45	34.6	495	12	Q89801	Q89801	cowpea mott
40	45	34.6	539	11	Q8C0J7	Q8c0j7	mus musculu
41	45	34.6	622	11	Q8CFM1	Q8cfm1	mus musculu
42	45	34.6	652	11	Q925U4	Q925u4	mus musculu
43	44.5	34.2	140	12	Q99A60	Q99a60	bovine vira
44	44.5	34.2	156	2	Q9X9Q5	Q9x9q5	sphingomona
45	44.5	34.2	261	13	Q9W6V1	Q9w6v1	gallus gall

# ALIGNMENTS

# RESULT 1

Q8P8D7

ID Q8P8D7 PRELIMINARY; PRT; 627 AA.  
AC Q8P8D7;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein XCC2305.  
GN XCC2305.  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=340;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33913 / NCPPB 528;  
RX MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities.";  
RL Nature 417:459-463(2002).  
DR EMBL; AE012338; AAM41584.1; -.  
DR InterPro; IPR005532; DUF323.  
DR Pfam; PF03781; DUF323; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 627 AA; 68307 MW; F3C311369D45CE74 CRC64;

Query Match 45.0%; Score 58.5; DB 16; Length 627;  
Best Local Similarity 50.0%; Pred. No. 1.5;  
Matches 12; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

Qy 1 WRCVLREGPAGGCAWFN---RHRL 21  
| | | | | | | | | |  
Db 566 WHASYRRAPADGAAWFNPGRQRL 589

# RESULT 2

Q9PFC6

ID Q9PFC6 PRELIMINARY; PRT; 621 AA.  
AC Q9PFC6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)





OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003528; AAF49542.2; -.  
 DR FlyBase; FBgn0036566; CG5284.  
 DR InterPro; IPR000644; CBS\_domain.  
 DR InterPro; IPR001807; Cl-channel\_volt.  
 DR Pfam; PF00571; CBS; 2.  
 DR Pfam; PF00654; voltage\_CLC; 1.  
 DR SMART; SM00116; CBS; 2.  
 SQ SEQUENCE 788 AA; 87259 MW; 995F6E8E5EE0177F CRC64;

Query Match 39.6%; Score 51.5; DB 5; Length 788;  
 Best Local Similarity 27.9%; Pred. No. 22;  
 Matches 12; Conservative 1; Mismatches 7; Indels 23; Gaps 1;

Qy 1 WRCVLREGPAGGCA-----WFNRHR 20  
 | | | | | :  
 Db 137 WLCVLLVGIAAGCVAGMVDIGASWMSDLKHGICPPAFWFNREQ 179

# RESULT 5

## Q8IQN2

ID Q8IQN2 PRELIMINARY; PRT; 822 AA.  
 AC Q8IQN2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG5284-PB.  
 GN CG5284.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";



RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003528; AAN11761.1; -.  
 SQ SEQUENCE 822 AA; 90547 MW; FA944F9FCCFFBF29 CRC64;

Query Match 39.6%; Score 51.5; DB 5; Length 822;  
 Best Local Similarity 27.9%; Pred. No. 23;  
 Matches 12; Conservative 1; Mismatches 7; Indels 23; Gaps 1;

Qy 1 WRCVLREGPAGGCA-----WFNRHR 20  
 | | | | | | | | | | :  
 Db 137 WLCVLLVGIAAGCVAGMVDIGASWMSDLKHGICPPAFWFNREQ 179

# RESULT 6

## Q8MQK2

ID Q8MQK2 PRELIMINARY; PRT; 882 AA.  
 AC Q8MQK2;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE LD07266p (Fragment).  
 GN CG5284.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY129438; AAM76180.1; -.

DR FlyBase; FBgn0036566; CG5284.  
 DR InterPro; IPR000644; CBS\_domain.  
 DR InterPro; IPR001807; Cl-channel\_volt.  
 DR Pfam; PF00571; CBS; 2.  
 DR Pfam; PF00654; voltage\_CLC; 1.  
 DR PRINTS; PR00762; CLCHANNEL.  
 DR SMART; SM00116; CBS; 2.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 882 AA; 96750 MW; 1B5BC76F34B0D24B CRC64;

Query Match 39.6%; Score 51.5; DB 5; Length 882;  
 Best Local Similarity 27.9%; Pred. No. 25;  
 Matches 12; Conservative 1; Mismatches 7; Indels 23; Gaps 1;

Qy 1 WRCVLREGPAGGCA-----WFNRHR 20  
 | | | | | :  
 Db 197 WLCVLLVGIAAGCVAGMVDIGASWMSDLKHGICPPAFWFNREQ 239

# RESULT 7

Q8GPQ9  
 ID Q8GPQ9 PRELIMINARY; PRT; 520 AA.  
 AC Q8GPQ9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SG17M;  
 RX MEDLINE=22313472; PubMed=12426355;  
 RA Larbig K.D., Christmann A., Johann A., Klockgether J., Hartsch T.,  
 RA Merkl R., Wiehlmann L., Fritz H.J., Tummeler B.;  
 RT "Gene Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity  
 RT on a Pseudomonas aeruginosa Clone.";  
 RL J. Bacteriol. 184:6665-6680(2002).  
 DR EMBL; AF440524; AAN62315.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 520 AA; 57774 MW; A51505FEFAA14F54 CRC64;

Query Match 38.5%; Score 50; DB 2; Length 520;  
 Best Local Similarity 38.1%; Pred. No. 24;  
 Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNRHRL 21  
 | | | | : | | : : :  
 Db 204 WRCFLQGLPIGRAPTFSKHQI 224

# RESULT 8

Q8X5L3  
 ID Q8X5L3 PRELIMINARY; PRT; 523 AA.  
 AC Q8X5L3;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative protease.  
 GN YHJS OR Z4952 OR ECS4416.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AE005580; AAG58679.1; -.  
 DR EMBL; AP002565; BAB37839.1; -.  
 DR InterPro; IPR002453; Beta\_tubulin.  
 DR PROSITE; PS00228; TUBULIN\_B\_AUTOREG; 1.  
 KW Protease; Complete proteome.  
 SQ SEQUENCE 523 AA; 59398 MW; 109DD02257AB1EF8 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 523;  
 Best Local Similarity 64.7%; Pred. No. 25;  
 Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 5 LREGPAGGCAWFN--RH 19  
 || |||| ||| ||  
 Db 20 LRHMPAGGVWVFNVD RH 36

# RESULT 9

## Q8FCH0

ID Q8FCH0 PRELIMINARY; PRT; 523 AA.  
 AC Q8FCH0;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein yhjS.  
 GN YHJS OR C4348.

OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 DR EMBL; AE016768; AAN82784.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 523 AA; 59298 MW; 3EFEA13A1512C504 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 523;  
 Best Local Similarity 64.7%; Pred. No. 25;  
 Matches 11; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 5 LREGPAGGCAWFN--RH 19  
 || |||| ||| ||  
 Db 20 LRHMPAGGVWVFNVDRL 36

# RESULT 10

Q9FBZ4

ID Q9FBZ4 PRELIMINARY; PRT; 1239 AA.  
 AC Q9FBZ4;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative secreted peptidase.  
 GN SCO7188 OR SC8A11.16C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete *Streptomyces*  
 RT *coelicolor* A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939130; CAC01588.1; -.  
 DR HSSP; Q99405; 1MPT.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS50840; PA; 1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1239 AA; 128505 MW; 8F5E9AC68EB1260A CRC64;

Query Match 38.5%; Score 50; DB 16; Length 1239;  
 Best Local Similarity 75.0%; Pred. No. 59;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGG 12  
 ||| |||||  
 Db 184 WRSVTGEGPAGG 195

# RESULT 11

Q8XT34

ID Q8XT34 PRELIMINARY; PRT; 245 AA.  
 AC Q8XT34;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein RSp0281.  
 GN RSP0281 OR RS03683.  
 OS *Ralstonia solanacearum* (*Pseudomonas solanacearum*).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; *Ralstonia*.  
 OX NCBI\_TaxID=305;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646077; CAD17432.1; -.  
 DR InterPro; IPR000160; GGDEF.  
 DR Pfam; PF00990; GGDEF; 1.  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 245 AA; 27122 MW; CEE9FCB0B6C86C27 CRC64;

Query Match 37.7%; Score 49; DB 16; Length 245;  
 Best Local Similarity 48.0%; Pred. No. 16;  
 Matches 12; Conservative 1; Mismatches 6; Indels 6; Gaps 1;

Qy 2 RCVLREG-----PAGGCAWFNRHR 20  
 | | : | | | | | | | |  
 Db 51 RAVIRCGNRRQHLFARKCAWRNRQR 75

# RESULT 12

## Q92ND1

ID Q92ND1 PRELIMINARY; PRT; 559 AA.  
 AC Q92ND1;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical transmembrane protein SMC01665.  
 GN R02277 OR SMC01665.  
 OS *Rhizobium meliloti* (*Sinorhizobium meliloti*).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; *Sinorhizobium*.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT *Sinorhizobium meliloti* strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL; AL591790; CAC46856.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 559 AA; 61358 MW; 1663C17276640F4F CRC64;

Query Match 37.7%; Score 49; DB 16; Length 559;  
 Best Local Similarity 45.0%; Pred. No. 37;

Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```
Qy      1 WRCVLREGPAGGCAWFNRHR 20
      || | : || || |:|
Db      415 WRSVTADRVAGSSAWLPRYR 434
```

RESULT 13

Q9HX92

ID Q9HX92 PRELIMINARY; PRT; 906 AA.  
AC Q9HX92;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Probable transcriptional regulator.  
GN PA3921.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PAO1;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
RT opportunistic pathogen."  
RL Nature 406:959-964(2000).  
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL; AE004809; AAG07308.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR000792; HTH\_LuxR.  
DR Pfam; PF00196; GerE; 1.  
DR PRINTS; PR00038; HTHLUXR.  
DR ProDom; PD000307; HTH\_LuxR; 1.  
DR SMART; SM00382; AAA; 1.  
DR SMART; SM00421; HTH\_LUXR; 1.  
DR PROSITE; PS00622; HTH\_LUXR\_FAMILY; 1.  
KW DNA-binding; Transcription regulation; Complete proteome.  
SQ SEQUENCE 906 AA; 101346 MW; CCC4FF2E0B414FC2 CRC64;

Query Match 37.7%; Score 49; DB 16; Length 906;  
Best Local Similarity 45.0%; Pred. No. 61;  
Matches 9; Conservative 2; Mismatches 3; Indels 6; Gaps 1;

```
Qy      8 GPAGG-----CAWFNRHRL 21
      ||: | | ||: || |
Db      356 GPSAGSLHLRACGWFSRHGL 375
```

RESULT 14

Q8X1W7

ID Q8X1W7 PRELIMINARY; PRT; 482 AA.  
AC Q8X1W7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Acid phosphatase.  
GN APH.  
OS Monascus anka.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Monascaceae; Monascus.  
OX NCBI\_TaxID=5098;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nagashima T., Anazawa H., Terasaki Y.;  
RT "nitrate reductase."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB046447; BAB84518.1; -.  
DR InterPro; IPR000560; HisAc\_phosphatase.  
DR Pfam; PF00328; acid\_phosphatase; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
SQ SEQUENCE 482 AA; 52779 MW; 6ADB89041D44D093 CRC64;

Query Match 36.9%; Score 48; DB 3; Length 482;  
Best Local Similarity 50.0%; Pred. No. 46;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 4 VLREGPAGGCAWFN 17  
:|:|:|:|  
Db 324 LLNQGPSAGTLWFN 337

RESULT 15

Q9X838

ID Q9X838 PRELIMINARY; PRT; 666 AA.  
AC Q9X838;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein SCO6072.  
GN SCO6072 OR SC9B1.19.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,



RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete *Streptomyces*  
RT *coelicolor* A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939126; CAB41565.1; -.  
DR InterPro; IPR002791; DUF89.  
DR Pfam; PF01937; DUF89; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 666 AA; 71463 MW; EF87B894A65E8B54 CRC64;

Query Match 36.9%; Score 48; DB 16; Length 666;  
Best Local Similarity 50.0%; Pred. No. 63;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGCAWFNR 18  
| | : | | | | : |  
Db 69 WGRVPLDRPARGCAWADR 86

Search completed: November 13, 2003, 09:51:11  
Job time : 57.3438 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 24.9375 Seconds  
 (without alignments)  
 35.630 Million cell updates/sec

Title: US-09-228-866-16  
 Perfect score: 130  
 Sequence: 1 WRCVLREGPAGGCAWFNRHRL 21

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	130	100.0	21	1 US-08-526-710-16	Sequence 16, Appl
2	130	100.0	21	3 US-08-862-855-16	Sequence 16, Appl
3	130	100.0	21	3 US-09-226-985-16	Sequence 16, Appl
4	130	100.0	21	4 US-09-227-906-16	Sequence 16, Appl
5	51.5	39.6	164	4 US-09-252-991A-21892	Sequence 21892, A
6	49.5	38.1	140	4 US-09-252-991A-25759	Sequence 25759, A
7	49	37.7	921	4 US-09-252-991A-20327	Sequence 20327, A
8	48.5	37.3	134	4 US-09-252-991A-30413	Sequence 30413, A
9	47	36.2	9	1 US-08-526-710-20	Sequence 20, Appl
10	47	36.2	9	3 US-08-862-855-20	Sequence 20, Appl
11	47	36.2	9	3 US-09-226-985-20	Sequence 20, Appl

12	47	36.2	9	4	US-09-227-906-20	Sequence 20, Appl
13	47	36.2	398	4	US-09-252-991A-26217	Sequence 26217, A
14	47	36.2	903	3	US-08-804-439A-22	Sequence 22, Appl
15	47	36.2	903	3	US-08-720-229-22	Sequence 22, Appl
16	46.5	35.8	128	4	US-09-461-325-465	Sequence 465, App
17	46.5	35.8	165	4	US-09-461-325-464	Sequence 464, App
18	46	35.4	43	2	US-08-488-161-42	Sequence 42, Appl
19	46	35.4	43	3	US-09-273-685-42	Sequence 42, Appl
20	46	35.4	43	5	PCT-US95-11934-42	Sequence 42, Appl
21	46	35.4	246	4	US-09-336-536-31	Sequence 31, Appl
22	46	35.4	341	4	US-09-336-536-29	Sequence 29, Appl
23	46	35.4	370	4	US-09-336-536-28	Sequence 28, Appl
24	46	35.4	714	4	US-09-308-345A-47	Sequence 47, Appl
25	45	34.6	181	4	US-09-252-991A-30203	Sequence 30203, A
26	45	34.6	518	4	US-09-252-991A-25967	Sequence 25967, A
27	45	34.6	882	4	US-09-252-991A-17653	Sequence 17653, A
28	44.5	34.2	904	6	5244792-4	Patent No. 5244792
29	44	33.8	108	2	US-08-598-873-6	Sequence 6, Appli
30	44	33.8	108	3	US-08-605-430-6	Sequence 6, Appli
31	44	33.8	139	4	US-09-252-991A-18984	Sequence 18984, A
32	44	33.8	298	4	US-09-252-991A-29045	Sequence 29045, A
33	44	33.8	423	1	US-08-445-746-2	Sequence 2, Appli
34	44	33.8	423	3	US-09-008-722-2	Sequence 2, Appli
35	44	33.8	587	4	US-09-252-991A-25368	Sequence 25368, A
36	44	33.8	832	4	US-09-252-991A-24866	Sequence 24866, A
37	44	33.8	950	4	US-09-449-285A-4	Sequence 4, Appli
38	43.5	33.5	141	4	US-09-252-991A-24137	Sequence 24137, A
39	43.5	33.5	677	4	US-09-252-991A-28529	Sequence 28529, A
40	43	33.1	79	4	US-09-252-991A-27207	Sequence 27207, A
41	43	33.1	165	4	US-09-252-991A-27759	Sequence 27759, A
42	43	33.1	558	4	US-09-252-991A-25673	Sequence 25673, A
43	43	33.1	660	4	US-09-462-606-57	Sequence 57, Appl
44	43	33.1	1011	4	US-09-252-991A-32419	Sequence 32419, A
45	43	33.1	1111	1	US-08-317-450B-15	Sequence 15, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-526-710-16

; Sequence 16, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-16

```

```

Query Match          100.0%; Score 130; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 WRCVLREGPAGGCAWFNRHRL 21
        ||||||||||||||||
Db      1 WRCVLREGPAGGCAWFNRHRL 21

```

## RESULT 2

```

US-08-862-855-16
; Sequence 16, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-16

```

```

Query Match          100.0%; Score 130; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 WRCVLREGPAGGCAWFNRHRL 21
        |||||
Db      1 WRCVLREGPAGGCAWFNRHRL 21

```

# RESULT 3

```

US-09-226-985-16
; Sequence 16, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-16

```

```

Query Match          100.0%; Score 130; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 WRCVLREGPAGGCAWFNRHRL 21
        ||||||||||||||||||||
Db      1 WRCVLREGPAGGCAWFNRHRL 21

```

#### RESULT 4

```

US-09-227-906-16
; Sequence 16, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-16

```

```

Query Match          100.0%; Score 130; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 WRCVLREGPAGGCAWFNRHRL 21
        |||
Db      1 WRCVLREGPAGGCAWFNRHRL 21

```

```

RESULT 5
US-09-252-991A-21892
; Sequence 21892, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21892
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21892

```

Query Match 39.6%; Score 51.5; DB 4; Length 164;  
Best Local Similarity 50.0%; Pred. No. 4.6;  
Matches 11; Conservative 0; Mismatches 8; Indels 3; Gaps 1;

Qy 1 WRC---VLREGPAGGCAWFNRH 19  
||| || || | ||  
Db 73 WRCRGRALRAGPRGRRRWPPRH 94

RESULT 6

US-09-252-991A-25759  
; Sequence 25759, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25759  
; LENGTH: 140  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25759

Query Match 38.1%; Score 49.5; DB 4; Length 140;  
Best Local Similarity 52.6%; Pred. No. 7.4;  
Matches 10; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy 1 WRCV-----LREGPAGGCA 14  
||| :| ||| |||  
Db 26 WRCARPGPGVRAGPALGCA 44

RESULT 7

US-09-252-991A-20327  
; Sequence 20327, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27



; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20327  
; LENGTH: 921  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20327

Query Match 37.7%; Score 49; DB 4; Length 921;  
Best Local Similarity 45.0%; Pred. No. 56;  
Matches 9; Conservative 2; Mismatches 3; Indels 6; Gaps 1;

Qy 8 GPAGG-----CAWFNRHRL 21  
||:| | ||:| |  
Db 371 GPSAGSLHLRACGWFSRHGL 390

RESULT 8

US-09-252-991A-30413  
; Sequence 30413, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30413  
; LENGTH: 134  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30413

Query Match 37.3%; Score 48.5; DB 4; Length 134;  
Best Local Similarity 39.4%; Pred. No. 9.8;  
Matches 13; Conservative 0; Mismatches 7; Indels 13; Gaps 2;

Qy 1 WRCVLREGPA-----GGCAW-----FNRHR 20  
|| || || | || || ||  
Db 6 WRTPLRRGPASAPRGHPRGDAAWTGRRSARRHR 38

RESULT 9

US-08-526-710-20  
; Sequence 20, Application US/08526710  
; Patent No. 5622699  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: Method of Identifying Molecules That  
; TITLE OF INVENTION: Home to a Selected Organ In Vivo

```

;   NUMBER OF SEQUENCES:  44
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Campbell and Flores
;     STREET:    4370 La Jolla Village Drive, Suite 700
;     CITY:      San Diego
;     STATE:     California
;     COUNTRY:   United States
;     ZIP:       92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:     IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:     PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/526,710
;     FILING DATE:        11-SEP-1995
;     CLASSIFICATION:     435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:               Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-LJ 1779
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:          (619) 535-9001
;     TELEFAX:            (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 9 amino acids
;       TYPE:   amino acid
;       TOPOLOGY: linear
;     MOLECULE TYPE:  peptide
US-08-526-710-20

```

```

Query Match          36.2%;  Score 47;  DB 1;  Length 9;
Best Local Similarity 100.0%;  Pred. No. 2.5e+05;
Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      4 VLREGPAGG 12
        |||||
Db      1 VLREGPAGG 9

```

```

RESULT 10
US-08-862-855-20
; Sequence 20, Application US/08862855
; Patent No. 6068829
;   GENERAL INFORMATION:
;     APPLICANT:  Ruoslahti, Erkki
;     APPLICANT:  Pasqualini, Renata
;     TITLE OF INVENTION:  Method of Identifying Molecules That
;     TITLE OF INVENTION:  Home to a Selected Organ In Vivo
;     NUMBER OF SEQUENCES:  44
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Campbell & Flores LLP
;       STREET:    4370 La Jolla Village Drive, Suite 700
;       CITY:      San Diego
;       STATE:     California
;       COUNTRY:   United States

```

```

;      ZIP: 92122
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/862,855
;      FILING DATE:
;      CLASSIFICATION: 424
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/526,710
;      FILING DATE: 11-SEP-1995
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/813,273
;      FILING DATE: 10-MAR-1997
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Campbell, Cathryn A.
;      REGISTRATION NUMBER: 31,815
;      REFERENCE/DOCKET NUMBER: P-LJ 2621
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (619) 535-9001
;      TELEFAX: (619) 535-8949
;      INFORMATION FOR SEQ ID NO: 20:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 9 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-862-855-20

```

```

Query Match          36.2%; Score 47; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      4 VLREGPAGG 12
        |||||
Db      1 VLREGPAGG 9

```

```

RESULT 11
US-09-226-985-20
; Sequence 20, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-20

```

```

Query Match          36.2%; Score 47; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      4 VLREGPAGG 12
        |||||
Db      1 VLREGPAGG 9

```

# RESULT 12

US-09-227-906-20

```

; Sequence 20, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States

```

```

;      ZIP: 92122
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/227,906
;      FILING DATE:
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/526,710
;      FILING DATE: 11-SEP-1995
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/813,273
;      FILING DATE: 10-MAR-1997
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/862,855
;      FILING DATE: 23-MAY-1997
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Campbell, Cathryn A.
;      REGISTRATION NUMBER: 31,815
;      REFERENCE/DOCKET NUMBER: P-LJ 3424
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (619) 535-9001
;      TELEFAX: (619) 535-8949
;      INFORMATION FOR SEQ ID NO: 20:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 9 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-09-227-906-20

```

```

Query Match          36.2%;  Score 47;  DB 4;  Length 9;
Best Local Similarity 100.0%;  Pred. No. 2.5e+05;
Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      4 VLREGPAGG 12
        |||||
Db      1 VLREGPAGG 9

```

```

RESULT 13
US-09-252-991A-26217
; Sequence 26217, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

```

; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 26217  
 ; LENGTH: 398  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-26217

Query Match 36.2%; Score 47; DB 4; Length 398;  
 Best Local Similarity 53.8%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WRCVLREGPAGGC 13  
 ||| | |  
 Db 59 WRCCCRRSPPKGC 71

RESULT 14

US-08-804-439A-22

; Sequence 22, Application US/08804439A

; Patent No. 6015565

; GENERAL INFORMATION:

; APPLICANT: Rose, Timothy M.

; APPLICANT: Bosch, Marnix L.

; APPLICANT: Strand, Kurt

; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV

; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES

; NUMBER OF SEQUENCES: 113

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 4225 Executive Square, Ste 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: USA

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,439A

; FILING DATE: February 21, 1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09176/004001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 678-5070

; TELEFAX: (619) 678-5099

; TELEX:

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 903 amino acids

; TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-804-439A-22
```

```
Query Match          36.2%; Score 47; DB 3; Length 903;
Best Local Similarity 58.3%; Pred. No. 1e+02;
Matches      7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      5 LREGPAGGCAWF 16
          :|:| | || ||
Db       1 MRQGAARGCRWF 12
```

RESULT 15

US-08-720-229-22

```
; Sequence 22, Application US/08720229
; Patent No. 6022542
```

; GENERAL INFORMATION:

```
; APPLICANT: Rose, Timothy M.
; APPLICANT: Bosch, Marnix L.
; APPLICANT: Strand, Kurt
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
```

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/08/720,229
; FILING DATE: 26-SEP-1996
; CLASSIFICATION: 424
```

; ATTORNEY/AGENT INFORMATION:

```
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20002.00
```

; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
```

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 903 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-08-720-229-22

Query Match 36.2%; Score 47; DB 3; Length 903;  
Best Local Similarity 58.3%; Pred. No. 1e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 5 LREGPAGGCAWF 16  
:|:| | | | |  
Db 1 MRQGAARGCRWF 12

Search completed: November 13, 2003, 09:55:00  
Job time : 25.9375 secs